

UNIVERSITE DE FRANCHE-COMTE

ECOLE DOCTORALE « LANGAGES, ESPACES, TEMPS, SOCIETES »

**Centre de recherche en linguistique et traitement automatique
des langues, Lucien Tesnière**

Thèse en vue de l'obtention du titre de docteur en

SCIENCES DU LANGAGE

**VERS UN PROTOTYPE DE TRADUCTION AUTOMATIQUE CONTROLEE
FRANÇAIS/ARABE APPLIQUEE AUX DOMAINES A SECURITE CRITIQUE**

Présentée et soutenue publiquement par

Mohand BEDDAR

Le 30 avril 2013

Sous la direction de Madame le Professeur Sylviane CARDEY-GREENFIELD

Membres du Jury :

Krzysztof BOGACKI, Professeur à l'université de Varsovie, Pologne, Rapporteur

Farouk BOUHADIBA, Professeur à l'université d'Oran, Algérie, Rapporteur

Sylviane CARDEY-GREENFIELD, Professeur à l'université de Franche-Comté

Christian FLUHR, Professeur à l'Institut Supérieur Arabe de Traduction (ISAT), Alger

Peter GREENFIELD, HDR, université de Franche-Comté

Laurent SPAGGIARI, Docteur en Linguistique, Airbus-France

Remerciements :

Je tiens à adresser mes remerciements les plus sincères :

tout d'abord à Madame le Professeur Sylviane CARDEY-GREENFIELD pour m'avoir guidé et soutenu, pour ses précieux conseils, pour ses encouragements et pour sa disponibilité lors de mes années de recherche.

à Monsieur Peter GREENFIELD pour ses conseils et ses remarques notamment pour toute la partie informatique.

à mes collègues de l'équipe du projet LiSe et à tous les chercheurs du Centre Lucien Tesnière avec qui j'ai eu des échanges constructifs tout au long de ma recherche.

à ma famille qui à travers son soutien permanent m'a encouragé et m'a soutenu tout au long de mes études, à ma femme Claudine qui m'a donné la force d'aller au bout de mes objectifs.

à mes amis et à toutes les personnes qui de près ou de loin, par leurs soutiens, par leurs conseils ou par leurs encouragements, ont contribué à mes travaux de recherche.

Sommaire :

Abréviations	15
---------------------------	-----------

Sommaire des figures	18
-----------------------------------	-----------

INTRODUCTION GENERALE

1. Contexte de la recherche	22
2. Objectifs de la recherche	24
3. Cadre théorique de la recherche	25
4. Organisation de la thèse	25

CHAPITRE 1 LANGUE SPECIALISEE

1. Introduction	29
2. Technique, scientifique ou de spécialité ?	30
3. Langue générale et langue de spécialité	31
4. Spécialité ou spécialisée	37
5. Spécialisée ou spécialisées	38
6. Degré de spécialisation (langue générale / langue commune / langue spécialisée)	39
7. Délimitation de la langue spécialisée	42
7.1. Spécialisation par domaine d'activité	42
7.2. Spécialisation par communauté de spécialistes	43
7.3. Spécialisation par traits linguistiques	44
8. D'un patrimoine langagier commun à son usage particulier	47
9. Langue contrôlée vs langue spécialisée	52
9.1. Diachronie / Synchronie	53
9.2. Supériorité / Antériorité	54
9.3. Langage / Métalangage	54
9.4. Ecrit / Oral	55
9.5. Monofonctionnel / Multifonctionnel	55

9.6. Monosémie / Polysémie	56
9.7. Lexique : limité / illimité	57
9.8. Règles linguistiques / Normes sociales	57
10. La langue spécialisée entre sémiotique et langue naturelle	58
11. Caractéristiques des discours spécialisés	59
12. Conclusion	62

CHAPITRE 2 PROTOCOLES DE SECURITE

1. La sécurité, un domaine sensible et hétéroclite	65
2. Corpus de travail.....	66
2.1. Protocoles	69
2.2. Messages d'alerte	70
2.2.1. Les Panneaux à Messages Variables, des messages contrôlés	71
2.2.2. Les alertes enlèvements	72
3. Domaines de travail	73
3.1. Sécurité civile	74
3.2. Médecine	75
3.3. Aéronautique	76
4. La communication écrite en milieu professionnel, un facteur de risque ?	78
5. Protocoles de sécurité, des textes procéduraux ?	79
6. Les textes procéduraux : plusieurs dénominations, des caractéristiques communes	82
7. Définition des protocoles de sécurité	85
8. Structure externe des protocoles de sécurité	85
8.1. Les introducteurs de cadres	86
8.2. Structure visuelle des protocoles de sécurité	89
8.3. Structures énumératives	90
8.4. Ensemble conditions/instructions	91
8.4.1. Structures de cas.....	92
8.4.1.1. Structures conditionnelles	92
8.4.1.2. Les titres	94
8.4.1.3. Les circonstanciels de but	94
8.4.1.4. Structures interrogatives	95

8.4.2. Structures procédurales	96
8.4.3. Structures déclaratives	97
9. Structure interne des protocoles de sécurité	99
9.1. Lexique ou terminologie	99
9.2. Protocoles de sécurité, une prédominance nominale	101
9.2.1. Morphologie lexicale	102
9.2.1.1. Lexique dérivé	104
9.2.1.1.1. Suffixation	105
9.2.1.1.2. Préfixation	108
9.2.1.2. Lexique composé	112
9.2.1.2.1. Composition savante	112
9.2.1.2.2. Composition syntagmatique	113
9.2.2. Unités syntagmatiques nominales	120
9.2.3. Combinatoire sémantique et abréviation syntagmatique.....	124
9.2.3.1. Combinatoire sémantique	124
9.2.3.2. Abréviation syntagmatique	126
9.3. Unités syntagmatiques verbales	127
9.3.1. Verbes support et TA	132
9.3.2. Relation verbe support/nom prédicatif	134
9.3.3. Pivot et paradigme	135
9.3.4. Critères d'identification des verbes support	137
9.3.5. Approche analytique et approche synthétique	140
9.3.6. Verbes support & verbes étendus	140
9.3.7. Verbes support et prépositions	144
9.3.8. Analyse systémique des constructions à support	146
9.3.8.1. Correspondances formes synthétiques/formes analytiques	146
9.3.8.2. Saisie des dépendances prépositions verbales/verbes support	148
9.4. L'infinitif	150
9.4.1. Les emplois de l'infinitif	151
10. La sécurité aux limites de la linguistique	154
11. Protocoles de sécurité et langages non linguistiques	155
12. Les dénominations spécialisées	157
13. Conclusion	158

CHAPITRE 3 LANGUE CONTROLEE EN VUE DE LA TRADUCTION AUTOMATIQUE VERS L'ARABE

I. LANGUES CONTROLEES, BREF APERÇU	161
1. Langues contrôlées, des besoins croissants	162
2. Quelques langues contrôlées existantes	163
3. Langue contrôlée LiSe, une réponse à des besoins multiples	165
 II. POURQUOI CONTROLER ?	167
1. Définition	168
2. S'affranchir des ambiguïtés par la langue contrôlée	169
2.1. Les ambiguïtés morphologiques	171
2.1.1. Descriptions des diverses formes d'ambiguïtés morphologiques	171
2.1.2. Solutions apportées par la langue contrôlée	172
2.2. Les ambiguïtés lexicales	172
2.2.1. Descriptions des diverses formes d'ambiguïtés lexicales	172
2.2.2. Solutions apportées par la langue contrôlée	174
2.3. Les ambiguïtés référentielles	174
2.3.1. Descriptions des diverses formes d'ambiguïtés référentielles	174
2.3.2. Solutions apportées par la langue contrôlée	176
2.4. Les ambiguïtés liées aux prépositions	177
2.4.1. Les prépositions, source de nombreuses ambiguïtés	177
2.4.2. Solutions apportées par la langue contrôlée	178
2.5. Les ambiguïtés syntaxiques	179
2.5.1. Descriptions des diverses formes d'ambiguïtés syntaxiques	179
2.5.2. Solutions apportées par la langue contrôlée	180
2.6. Mise en page, clarté et homogénéité	180
2.6.1. Les problématiques liées à la mise en page et à la clarté	180
2.6.2. Les propositions et pistes d'amélioration concernant la mise en page	182
3. Conclusion	183
 III. LANGUE CONTROLEE ET ANALYSE COMPARATIVE	184
1. Introduction	185
2. Langue contrôlée et TA	186

2.1. Langue contrôlée LiSe et système TACCT	186
2.2. Contrôle par analyse comparative	187
2.2.1. Contrôle statique	188
2.2.2. Contrôle dynamique	189
2.2.2.1. Ambiguïtés lexicales	190
2.2.2.2. Ambiguïtés syntaxiques	192
2.2.2.3. Divergences pragmatiques	193
2.3. Langue contrôlée LiSe et nominalisation	194
2.3.1. Nominalisation et langue contrôlée LiSe	194
2.3.2. La nominalisation outil de contrôle	195
2.3.3. La nominalisation outil normatif et génératif de structures	196
2.3.4. Contrôle bilingue par nominalisation	199
3. Conclusion	200

CHAPITRE 4 SYSTEME LINGUISTIQUE ARABE

1. Introduction	203
2. La phonétique dans la langue arabe	204
2.1. L'alphabet et caractères spéciaux	204
2.2. La prononciation et l'écriture de chaque lettre de l'alphabet	206
3. Morphologie de la langue arabe	207
3.1. Le nom et le nom-adjectif.....	207
3.1.1. Le genre, généralités	207
3.1.1.1. Genre des êtres animés sexués	207
3.1.1.2. Genre des objets asexués	208
3.1.1.3. Genre des collectifs	209
3.1.1.4. Genre dans les noms-adjectifs	209
3.1.2. Le nombre dans les noms et les noms-adjectifs	210
3.1.2.1. Le pluriel externe	210
3.1.2.2. Le pluriel interne	211
3.2. L'adjectif qualificatif	211
3.2.1. Accord de l'adjectif épithète	212
3.2.2. Accord de l'adjectif attribut	213
3.3. Les nombres	213
3.3.1. Les cardinaux	214
3.3.2. Les ordinaux	218

3.4. Flexion	220
3.4.1. Flexion à trois cas	220
3.4.2. Flexion à deux cas	221
3.4.2.1. Diptotes	221
3.4.2.2. Flexion du duel	222
3.4.2.3. Flexion du pluriel externe masculin	223
3.4.2.4. Flexion du pluriel externe féminin	223
3.4.2.5. Flexion des mots à finale َ ou ي.....	224
3.4.3. Mots sans flexion	224
3.4.4. Emploi des cas	225
3.4.4.1. Le cas nominatif	225
3.4.4.2. Le cas accusatif	225
3.4.4.3. Le cas génitif	226
3.5. Le verbe	226
3.5.1. Accompli	227
3.5.2. Inaccompli	228
3.5.2.1. Inaccompli indicatif	229
3.5.2.2. Inaccompli subjonctif	230
3.5.2.3. Inaccompli apocopé	231
3.5.3. Impératif	232
3.6. Les prépositions	233
3.6.1. Généralités	233
3.6.2. La préposition ب	234
3.6.3. La préposition في	234
3.6.4. La préposition مِنْ	235
3.6.5. La préposition عَنْ	235
3.6.6. La préposition إِلَى	236
3.6.7. La préposition لِ	236
3.6.8. La préposition عَلَى	236
3.6.9. La préposition كَ	237
3.6.10. La préposition حَتَّى	237
3.6.11. La préposition مَعَ	237
3.6.12. La préposition عِنْدَ	238
3.7. Les pronoms	238
3.7.1 Pronoms personnels isolés	238
3.7.2 Pronoms personnels affixes	239

4. Syntaxe de la langue arabe	241
4.1. Les phrases nominales	241
4.1.1. Valeur et structure	241
4.1.2. Accord de l'attribut	242
4.1.3. Les complexes à l'intérieur du syntagme	243
4.1.3.1. L'apposition	243
4.1.3.2. L'utilisation des démonstratifs	244
4.1.3.3. La détermination et l'indétermination	245
4.1.3.3.1. Indétermination	245
4.1.3.3.2. Détermination	245
4.1.3.3.2.1. Détermination par l'article ال.....	246
4.1.3.3.2.2. Détermination par annexion	246
4.1.3.4. L'élatif	247
4.1.3.5. Le superlatif	247
4.2. Les phrases verbales	248
4.2.1. Valeur et structure	248
4.2.2. Accord du verbe	248
4.3. Ordre des mots	249
5. Conclusion	250

CHAPITRE 5 TRADUCTION AUTOMATIQUE

1. Introduction	253
2. Histoire de la traduction automatique	253
2.1. De l'engouement des débuts à la désillusion du rapport ALPAC	253
2.2. De 1965 à 1975 : un domaine en berne	255
2.3. De 1975 à nos jours : le renouveau	255
3. Traduction automatique et traduction assistée	257
4. Les différentes architectures des systèmes de traduction automatique.....	258
4.1. Approche directe et approche indirecte.....	258
4.2. Approche par pivot et approche par transfert.....	258
4.2.1. Approche par pivot	259
4.2.2. Approche à transfert	259
5. Conclusion	260

CHAPITRE 6 SYSTEMIQUE ET SYSTEME

1. Introduction	263
2. Naissance de la systémique	263
3. Voyage dans le temps, bref historique	264
4. Raisons d'une apparition	268
5. Définition d'un système	269
5.1. La complexité	270
5.2. La globalité	272
5.3. L'interaction	273
5.4. Autres concepts opérationnels	274
6. Systémique linguistique	276
6.1. Théorie SyGULAC	276
6.2. Méthodologie	277
7. Difficultés de traduire automatiquement	278
7.1. Difficultés de modélisation	278
7.2. Difficultés de liaison	279
8. Systémique et système de traduction TACCT	280
8.1. Complexité	280
8.2. Globalité	281
8.3. Interaction	283
9. Axe analytique et axe classificatoire	284
10. Conclusion	285

CHAPITRE 7 MODELE SYNTAXICO-SEMANTIQUE DU SYSTEME TACCT

1. Introduction	287
2. Notion de macrostructure miroir	288
3. Modèle syntaxico-sémantique d'un point de vue systémique	290
3.1. Axe analytique	290
3.1.1. Groupe verbal /macrostructure	291
3.1.2. Microstructures	292

3.2. Axe classificatoire	293
4. Les prépositions liées aux verbes dans le modèle syntaxico-sémantique TACCT	295
5. Interaction lexicale / syntaxe	297
5.1. L'opérateur lexis	298
5.1.1. Opérateur lexis dans les macrostructures	298
5.1.2. Opérateur lexis dans les microstructures	299
6. Groupes verbaux du français contrôlé	300
6.1. Verbes à l'infinitif	300
6.2. Verbes conjugués	302
7. Structures équivalentes en arabe	303
7.1. Verbes infinitifs en arabe	305
7.2. Verbes conjugués en arabe	306
8. Transfert des macrostructures	307
8.1. Transfert frC/frC	307
8.2. Transfert frC/arC	309
9. Éléments constitutifs des structures élémentaires	312
9.1. Verbes infinitifs et conjugués	313
9.1.1. Transfert des verbes infinitifs	314
9.1.1.1. Le masdar équivalent de l'infinitif à valeur impérative	315
9.1.1.2. Le masdar sujet du verbe impersonnel (يجب)	316
9.1.2. Transfert des verbes conjugués	317
9.1.2.1. Les conjonctions <i>si /quand/dès que</i>	318
9.1.2.2. La négation	318
9.1.2.3. Genre de l'arg0	320
9.1.2.3.1. Sujet lexical	321
9.1.2.3.2. Sujet pronom <i>vous</i>	322
9.1.2.3.2.1. L'arabe langue pro-drop	322
9.1.2.3.2.2. Arg 0 un sujet à valeur nulle	323
9.2. Les arguments	327
9.2.1. Propriétés sémantiques des arguments	328
9.2.2. Transfert des arguments	329
9.2.3. Casualité distributionnelle	331
9.2.3.1. Cas généraux	331
9.2.3.2. Les sous-cas	333

9.2.4. Accord nom/adjectif	336
9.2.4.1. Principe de super-catégorisation	338
9.3. Particule verbale (part_v)	341
9.4. Attribut (att)	343
9.4.1. Le transfert des structures attributives	344
9.4.1.1. Structures attributives en vinf	344
9.4.1.1.1. Cas de l'attribut du sujet	344
9.4.1.1.2. Cas de l'attribut de l'arg1	345
9.4.1.2. Structures attributives en si	346
9.4.1.2.1. Cas de l'attribut du sujet	346
9.4.1.2.2. Cas de l'attribut de l'arg1	347
10. Eléments facultatifs des macrostructures contrôlées	348
10.1. La négation	348
10.1.1. Dans les structures en vinf	348
10.1.1.1. Neg2 = <i>pas</i>	349
10.1.1.2. Neg2 = <i>jamais</i>	349
10.1.1.3. Conclusion sur la négation dans les structures en vinf	350
10.1.2. Dans les structures en si	350
10.1.2.1. La conjonction est : إذا / <i>si</i>	351
10.1.2.2. La conjonction est : <i>quand</i> / عندما , <i>dès que</i> / بمجرد أن	351
10.1.2.3. Conclusion sur la négation dans les structures en si	352
10.2. Les compléments	353
10.2.1. Nature des compléments	354
10.2.1.1. Les compléments de temps	355
10.2.1.2. Les compléments de lieu	356
10.2.1.3. Les compléments de manière	356
10.2.1.4. Les compléments de moyen	356
10.2.1.5. Les compléments d'instrument	357
10.2.1.6. Les autres compléments	357
10.2.2. Tableau synthétique des prépositions.....	358
11. Traduction automatique et syntagmes nominaux.....	359
11.1. Syntagmes nominaux N1-de-N2	360
11.1.1. Syntagmes nominaux libres N1-de-N2.....	361
11.1.1.1. Règles de transfert des syntagmes nominaux libres N1-de-N2.....	362
11.1.1.2. Représentation à séquences terminales finie et infinie	366
11.1.1.2.1. Représentation isomorphique infinie à séquences terminales	366
11.1.1.2.2. Représentation isomorphique finie à séquences terminales	367

11.1.1.3. Isomorphie à séquences terminales hiérarchique	368
11.1.2. Syntagmes lexicaux N1- <i>de</i> -N2	372
11.1.3. Syntagmes lexicaux libres et Système TACCT français-arabe	373
11.2. Les groupes nominaux N1- <i>adj-de</i> -N2	376
12. Avantages du système TACCT	378
12.1. Synchronisation phase d'analyse et phase de transfert	378
12.2. Maîtrise des diathèses	381
12.3. Désambiguïsation des relations syntaxiques	383
12.4. Facilité d'accès aux données	385
12.5. Traçabilité, extensibilité	386
13. Etapes du transfert et architecture générale	387
13.1. Les différentes étapes du transfert	387
13.2. Architecture générale du système	388
14. Structuration des données	389
14.1. Description technique de l'interface LiSe	391
14.2. Choix de base pour la sélection des bonnes plages de données linguistiques	391
14.3. Choix des structures syntaxiques par le rédacteur	393
14.3.1. Choix du type de phrase	393
14.3.2. Niveaux d'analyses des différentes structures	394
15. Conclusion.....	396

CONCLUSION GENERALE398

REFERENCES ET BIBLIOGRAPHIE

1. Références.....	402
2. Bibliographie.....	410
3. Sites internet.....	421

ANNEXES

1. Corpus Virus H1N1 contrôlé.....	424
2. Corpus Virus H1N1 contrôlé traduit.....	425

3. Corpus Régulation contrôlé.....	426
4. Corpus Régulation contrôlé traduit.....	427
5. Corpus Désobstruction non contrôlé.....	428
6. Corpus Désobstruction contrôlé.....	429
7. Corpus Tunnel Mont Blanc non contrôlé.....	432
8. Corpus Tunnel Mont Blanc contrôlé.....	433
9. Corpus Hémorragie non contrôlé.....	435
10. Corpus Hémorragie contrôlé.....	436

Abréviations :

***** : possibilité non envisageable

****** : exemple non issu du corpus

acc : accusatif

accdef : accusatif déficient

ad : article défini

adj : adjectif

adjfd : adjectif féminin duel

adjfp : adjectif féminin pluriel

adjfs : adjectif féminin singulier

adjmd : adjectif masculin duel

adjmp : adjectif masculin pluriel

adjms : adjectif masculin singulier

adv : adverbe

adv.interr : adverbe interrogatif

ai : article indéfini

ar : arabe

arC : arabe contrôlé

arg1 : argument 1

arg2 : argument 2

artu : article unique (الـ)

att : attribut

comp1 : complément 1

comp2 : complément 2

det : déterminant

detv : déterminant variable

fr : français

frC : français contrôlé

GN : groupe nominal

gn_sd : groupe nominal sans déterminant

LC : Langue cible
LiSe : Linguistique et Sécurité
LS : langue source
n : nom
N1 : nom 1
N2 : nom 2
neg1 : négation 1
neg2 : négation 2
nfd : nom féminin duel
nfp : nom féminin pluriel
nfs : nom féminin singulier
nmd : nom masculin duel
nmp : nom masculin pluriel
nms : nom masculin singulier
no : nominatif
nodef : nominatif déficient
Npred : nom prédicatif
num : numéral
nv : nom du verbe
nver : nom verbal
opt : optionnel
P : phrase
P_{act} : participe actif
prep : préposition
prep_comp : préposition introduisant un complément
prep_v : préposition verbale
pron : pronom
pt : point
qqc : quelque chose
qqn : quelqu'un
qqp : quelque part
SN : syntagme nominal

suf : suffixe

SV : syntagme verbal

t : tanwin

TA : traduction automatique

TACCT : Traduction Automatique Contrôlée Centre Tesnière

u_{quan} : unité de quantification

v : verbe

vconj : verbe conjugué

v_{inacc} : verbe à l'inaccompli

vinf : verbe infinitif

Vsup : verbe support

Sommaire des figures :

Figure 1: Langue générale et langue spécialisée	36
Figure 2: Rapport spécialisation/nombre de locuteurs	41
Figure 3: Langue générale, langue commune et langue spécialisée	46
Figure 4: Verbe simple et verbe support	128
Figure 5 : Contrôle statique et contrôle dynamique	190
Figure 6 : Les contraintes lexicales facilitent les équivalences référentielles	192
Figure 7 : Contrôle LiSe et nominalisation	198
Figure 8 : Contrôle bilingue par nominalisation	200
Figure 9 : Systémique appliquée à la traduction automatique	284
Figure 10 : Notion de macrostructure miroir dans le système TACCT	289
Figure 11 : Exemple d'isomorphisme par paire (groupe verbal/macrostructure) entre le français et l'arabe	292
Figure 12 : L'analyse lexico-syntaxique	297
Figure 13 : Module frC_groupes recensant les macrostructures du français contrôlé ...	307
Figure 14 : Modules d'analyse du système TACCT	308
Figure 15 : Trace LPA Prolog de la phase d'analyse	309
Figure 16 : Modules de transfert du système TACCT	310
Figure 17 : Module arC_groupes recensant les macrostructures de l'arabe contrôlé ...	310
Figure 18 : Trace LPA Prolog de la phase de transfert	312
Figure 19 : Eléments constitutifs des macrostructures	313
Figure 20 : Module de transfert des groupes verbaux	314
Figure 21 : Noms verbaux composés	316

Figure 22 : Règles de transfert des verbes conjugués	324
Figure 23 : Règles de transfert de la négation	325
Figure 24 : Transfert des syntagmes nominaux de type (ad+n+adj)	326
Figure 25 : Transfert des verbes conjugués composés	326
Figure 26 : Propriétés sémantiques des arguments	329
Figure 27 : Distribution des cas accusatif et nominatif sur les arguments	330
Figure 28 : Marquage casuel général des arguments	331
Figure 29 : Cas généraux des compléments	332
Figure 30 : Unités lexicales non déclinées dans le dictionnaire lexical	334
Figure 31 : Casualité distributionnelle du système TACCT	335
Figure 32 : Règles d'accord des adjectifs en langue arabe	337
Figure 33 : Trace LPA Prolog des accords noms/adjectifs pendant la phase de transfert	338
Figure 34 : Principe de super-catégorisation lors de la phase d'analyse du système TACCT	340
Figure 35 : Transfert des attributs	347
Figure 36 : Règles de transfert de la neg1 et neg2	353
Figure 37 : Système de détermination des syntagmes nominaux libres (N1-de-N2) en arabe	362
Figure 38 : Traduction des syntagmes nominaux libres (N1-de-N2) du français vers l'arabe. Cas du N1 contenant et N2 contenance	365
Figure 39 : Représentation isomorphique à séquences terminales des groupes nominaux N1-de-N2 montrant le degré de liberté de leur traduction vers l'arabe	367

Figure 40 : Relation d'inclusivité entre deux représentations isomorphiques à séquences terminales	370
Figure 41 : Relation d'exclusivité entre deux représentations isomorphiques à séquences terminales	371
Figures 42 : Syntagmes nominaux lexicaux libres N1- <i>de</i> -N2 en arabe	372
Figure 43 : Traduction des syntagmes nominaux lexicaux libres (N1- <i>de</i> -N2) du français vers l'arabe	375
Figure 44 : Synchronisation phase d'analyse et phase de transfert dans le système TACCT	380
Figure 45 : Analyse syntaxico-grammaticale du verbe « <i>rendre</i> »	384
Figure 46 : Organisation des données dans le système TACCT	386

INTRODUCTION

GENERALE

1. Contexte de la recherche

La présente recherche porte sur la traduction automatique contrôlée français-arabe dans les domaines à sécurité critique. C'est un travail de recherche transverse qui traite à la fois des langues contrôlées qui font de plus en plus d'adeptes dans les communautés de spécialistes et la traduction automatique français-arabe. Elle est une réponse concrète à un constat indéniable qui replace la communication dans les domaines hautement sensibles comme celui de la sécurité au cœur des préoccupations majeures des gouvernements et des sociétés civiles et qui soulève nombre de questions autour du rôle que doivent jouer les langues naturelles comme l'un des moyens les plus élaborés de communication dans les situations de crise.

Nul ne peut contester aujourd'hui la place prépondérante qu'occupe la sécurité dans le quotidien des personnes et les enjeux qu'elle représente au sein des sociétés modernes. Ces sociétés davantage complexes et interconnectées manifestent une vulnérabilité flagrante qui les oblige à repenser leurs moyens d'organisation et de protection. Incendies, inondations, séismes, tsunamis, crashes d'avion, épidémies, accidents nucléaires, actes de malveillance, erreurs médicales, il ne passe pas un jour sans qu'un de ces sinistres fasse la une de nos journaux avec des bilans en pertes humaines et matérielles considérables. Cela sans mentionner les différents accidents de moindre importance qui n'ont pas été médiatisés tels que les accidents domestiques quotidiens mais qui peuvent néanmoins avoir des conséquences lourdes sur tous les plans.

Il est donc établi aujourd'hui que le risque zéro n'existe pas et que les sinistres font malheureusement partie de notre quotidien. Même si certains, par méconnaissance, considèrent que les sinistres sont une particularité d'un pays, d'une région, ou d'un endroit dans le monde, force est de constater que les risques aujourd'hui sont planétaires et qu'aucun endroit au monde ne peut être épargné et même quand ces risques semblent locaux ils peuvent avoir des répercussions planétaires, en témoigne la dernière catastrophe nucléaire survenue dans la province de Fukushima au Japon. Tous les pays et toutes les régions sont donc exposés à des risques de tout genre qui peuvent

mettre en péril leur développement et menacer la sécurité et la qualité de vie de leurs populations. Dans ce contexte, il est impératif de mettre en place des moyens efficaces pour gérer les situations de crise et notamment celui de la communication, important à la fois pour prévenir mais aussi pour gérer en temps réel une situation d'urgence.

La communication joue un rôle majeur dans le transfert des connaissances et la transmission de l'information. Elle permet aujourd'hui, par des moyens informatisés, de mettre en contact un ensemble d'individus ou une communauté de spécialistes et de maintenir le flux informationnel. Dans un monde de plus en plus mondialisé avec une libre circulation des biens et des personnes et où des milliers de langues cohabitent, l'enjeu de la communication est double car il n'est pas seulement question de transmettre un message dans une seule langue mais dans plusieurs, soulevant ainsi la question du multilinguisme.

La communication écrite à l'aide de systèmes informatisés est l'une des formes de communication la plus souvent utilisée pour le transfert des connaissances nécessaires à l'accomplissement des tâches et le déroulement des diverses actions. Elle donne une cohérence et une traçabilité dans la transmission des messages surtout quand ces derniers revêtent une fonction de procédure. Toutefois, et contrairement à une idée bien ancrée qui tend à associer les risques d'une mauvaise communication uniquement à l'oral, l'usage de la langue écrite peut lui aussi comporter des risques. Ils sont d'ailleurs bien plus importants car le rédacteur et le récepteur ne sont pas forcément dans une relation directe qui permettrait d'éclaircir certains malentendus. En effet des messages mal écrits peuvent conduire à de réelles catastrophes et à des conséquences irréversibles notamment dans des domaines jugés sensibles tels que les domaines à sécurité critique. La langue naturelle est un système complexe qui regorge de subtilités et d'ambiguïtés à tous les niveaux : lexical, syntaxique, sémantique et phonétique. Son usage doit être maîtrisé et obéir à des normes pour éviter toute erreur ou malentendu dans la transmission des messages qu'elle soit monolingue ou multilingue. Notre travail de recherche s'inscrit dans un vaste projet, intitulé LiSe (Linguistique et Sécurité) conduit par le Centre Lucien Tesnière. Ce projet est une approche novatrice dans les langues contrôlées et la traduction automatique appliquées aux domaines à sécurité

critique. L'analyse microsystemique de la langue avec un travail en intension sur le corpus recherche toutes les subtilités des langues et permet de définir des normes pour la rédaction de protocoles de sécurité et d'alertes ainsi que leur traduction automatique vers l'arabe et vers d'autres langues (anglais, chinois et thaï).

2. Objectifs de la recherche

Le présent travail de recherche vise la création d'un système de traduction automatique contrôlée à base de règles linguistiques, du français vers l'arabe, appliquée au domaine de la sécurité. Sous cet objectif principal, les motivations de ce travail de recherche sont multiples :

- Comprendre, sous tous les aspects discursifs internes et externes, le fonctionnement des textes de sécurité (protocoles de sécurité et alertes) ;
- Mettre en lumière les principales divergences linguistiques liées au fonctionnement de chacune des deux langues, le français et l'arabe, en s'appuyant sur un travail sur le corpus en intension ;
- Établir des normes de rédaction des protocoles de sécurité et des alertes qui permettent d'augmenter leur compréhensibilité et leur lisibilité ainsi que leur traduction vers l'arabe ;
- Mettre en place des moyens normatifs de contrôle dans deux perspectives différentes, une perspective intralange qui concernerait chacune des deux langues séparément et une perspective interlangue qui concernerait les deux langues simultanément, le français comme langue source et l'arabe comme langue cible ;
- Mettre en place, en s'appuyant sur les conclusions de l'étude comparative entre le français et l'arabe, un modèle syntaxico-sémantique isomorphe qui servirait de base au système de traduction automatique à base de règles linguistiques, TACCT.

3. Cadre théorique de la recherche

Pour les besoins de notre recherche et de par la finesse des analyses que nous voulons atteindre, que ce soit pour le contrôle ou pour la traduction automatique français-arabe, nous estimons que l'approche la plus adaptée à nos objectifs est l'approche microsystemique à travers la théorie SyGULAC (Systemic Grammar Using a Linguistically motivated Algebra and Calculus) élaborée au Centre Lucien Tesnière. L'analyse linguistique microsystemique du Centre Lucien Tesnière propose l'idée que, pour être traitées en toute sécurité, les langues doivent être décomposées en microsystemes assez petits mais aussi complets, analysables par un être humain et une machine. En plus de cela, les systemes ainsi délimités peuvent avoir des interactions avec d'autres systemes, et cette interaction est une propriété du langage. Rien n'est indépendant : le lexique, la morphologie, la syntaxe sont liés. L'avantage de la théorie SyGULAC déjà utilisée avec des résultats remarquables dans plusieurs applications telles que le sense data mining, les systemes de questions-réponses, etc., est la finesse et la profondeur des analyses qu'elle apporte. Le modèle syntaxico-sémantique du système TACCT (Traduction Automatique Contrôlée Centre Tesnière) français-arabe repose sur une analyse microsystemique intra- et interlangue assurée par la théorie SyGULAC. Elle permet de rapprocher les deux systemes linguistiques français et arabe et d'atteindre un isomorphisme dans la modélisation des structures appelées macrostructures miroir.

4. Organisation de la thèse

Le premier chapitre est consacré au concept des langues spécialisées. La démarche de définir la langue spécialisée nous semble pertinente avant toute tentative de formalisation car elle contribue à évaluer le degré de spécialisation du domaine sur lequel nous travaillons. Le concept des langues spécialisées sera aussi confronté, dans ce chapitre à celui des langues contrôlées, chères à la communauté industrielle, afin de saisir la nature de chacune des deux entités, comprendre les rapports qui les lient et connaître les points de jonction et de disjonction.

Le deuxième chapitre porte sur la nature des textes que nous traitons en l'occurrence les protocoles de sécurité et les alertes ainsi que sur les domaines auxquels ils appartiennent. Un grand nombre de corpus a été constitué dans les différents domaines traités : sécurité civile, médecine et aéronautique. Plus de 400 protocoles de sécurité monolingues et/ou bilingues ont été analysés. Ce chapitre fournit une étude détaillée des particularités linguistiques de ces textes ainsi que de leurs structures interne et externe.

Le troisième chapitre divisé en trois sous-chapitres, aborde la langue contrôlée. Il donne en premier lieu un bref historique des langues contrôlées et une description de certaines langues contrôlées déjà existantes. En deuxième lieu, ce chapitre à travers les deux derniers sous-chapitres donne les pistes de contrôle en vue de la traduction automatique du français vers l'arabe et définit les différents moyens normatifs de contrôle.

Le quatrième chapitre se penche sur le système linguistique de l'arabe, il passe en revue les principales caractéristiques du fonctionnement de la langue arabe. L'approche utilisée pour décrire le système linguistique de la langue arabe est une approche microsystemique. Chaque microsystème est décrit en prenant en compte les liens qu'il entretient avec les autres microsystèmes.

Le cinquième chapitre revient sur la théorie microsystemique et sur les conditions dans lesquelles cette théorie a vu le jour. Il aborde ensuite les premiers travaux en linguistique systemique réalisés par le Centre Lucien Tesnière à travers la théorie SyGULAC et propose une adaptation de cette dernière à la traduction automatique contrôlée français-arabe.

Le sixième chapitre présente un état de l'art et un retour sur les différentes approches adoptées pour la création des systèmes de traduction automatique.

Le septième et le dernier chapitre présente en détail le modèle syntaxico-sémantique du système TACCT français-arabe. Il décrit minutieusement le processus de création de ce modèle linguistique isomorphe qui permet des traductions fiables et

rapides du français vers l'arabe dans le domaine de la sécurité. Il présente également le concept des macrostructures-miroir en décrivant l'ensemble des éléments les constituant. Il décrit enfin l'architecture fonctionnelle du système TACCT, le processus de traduction avec toutes les étapes ainsi que l'interface utilisateur.

Chapitre 1

LANGUE SPECIALISÉE

1. Introduction

Le système de traduction automatique sur lequel nous travaillons est un système axé sur un domaine précis, celui de la sécurité. Il est donc important de comprendre le mode de fonctionnement des textes qui traitent de la sécurité et de saisir les principaux aspects linguistiques qui les caractérisent. Chaque domaine de spécialité a sa façon de formuler ses connaissances, et la langue en tant qu'instrument d'expression et de communication prend des formes variées pour les véhiculer. C'est là tout l'intérêt de ce chapitre, comprendre comment la langue s'adapte pour mieux communiquer et quelles sont les variations qui l'affectent.

Nul ne peut nier l'importance de la langue de spécialité dans le monde moderne et l'intérêt qu'elle suscite au sein de la communauté linguistique et industrielle. Personne n'échappe aujourd'hui à son essor et nombreux sont ceux qui y portent un intérêt particulier : les terminologues, les lexicographes, les enseignants de la langue de spécialité, les traducteurs et les interprètes spécialisés. Les concepteurs de traducteurs automatiques en font partie, conscients que toute tentative de formalisation doit passer par une connaissance approfondie de la langue en question, sont les premiers à exprimer cet engouement à étudier la langue de spécialité et à connaître son fonctionnement. C'est pour cela que nous estimons qu'avant de se lancer dans une quelconque démarche d'automatisation de la langue, il est nécessaire de se pencher sur son analyse non seulement parce que s'aventurer sur un terrain méconnu porte des risques d'échec mais aussi parce que la langue de spécialité attire et passionne de part sa richesse et sa variété.

Nous verrons au cours de ce chapitre que la langue de spécialité se démarque de la langue générale par des particularités à plusieurs niveaux : lexical et graphique, textuel et syntaxique et par un mode de fonctionnement spécifique. Ces particularités sont dues principalement à une récurrence de traits linguistiques que nous pourrions qualifier d'habitudes linguistiques. Nous tenterons ensuite d'appliquer ces constats sur les protocoles de sécurité afin d'estimer le degré de spécialisation de nos textes et de dégager leurs traits communs.

4. Spécialité ou spécialisée

Le débat sur la langue des domaines techniques a été aussi celui de la dénomination, langue de spécialité pour certains, langue spécialisée ou technolecte pour d'autres, les appellations n'ont cessé de fluctuer : langues spéciales (F. Saussure, 1975), langue spéciale et technique (J. Vendryès, 1968), langue de groupes particuliers (A. Meillet, 1975), langue des sciences (M. Fuchs, 1966), langues techniques (A. Dauzat, 1967), langue professionnelle (W. Wartburg, 1963) ou encore langues techniques et scientifiques (B. Quemada, 1955).

La finesse de ce débat a été telle que les chercheurs ont été amenés à reconsidérer la désinence du pluriel souvent rattachée à ces différentes dénominations (langues de spécialité, langues spécialisées, technolectes, etc.). (Voir chap1/5. Spécialisée ou spécialisées)

« Ajoutons, enfin, une remarque relative à l'opposition singulier/pluriel du terme langue : faut-il dire langue ou langues de spécialités... » (Kocourek, 1991)

En anglais, on utilise surtout Language for Special Purposes (LSP), Language for Specific Purposes et Special Languages comme des synonymes.

Comme nous l'avons abordé précédemment, et dans le souci de donner une progression logique à nos idées, nous préférons adopter dorénavant la dénomination langue spécialisée pour bon nombre de raisons. L'utilisation du participe passé souligne d'une part la porosité de la langue spécialisée et confirme encore une fois qu'il ne s'agit pas d'un ensemble « clos et imperméable » et qu'elle est en interaction permanente avec la langue commune à l'intérieur du système unique de la langue générale³

« Chacune des langues, langue technique, langue scientifique, langue commune, forme ainsi un ensemble qui n'est ni clos ni imperméable. Il y a au contraire une interaction constante

³ « Une langue donnée est donc constituée par un ensemble diversifié de sous-codes que le locuteur emploie en fonction de ses besoins d'expression et selon les caractéristiques de chaque situation de communication. Cependant, au delà de cette diversité foisonnante, toute langue possède un ensemble d'unités et de règles que tous ses locuteurs connaissent » (Cabré, 1998)

entre elles, qui forme une dynamique permanente et en perpétuel renouvellement».
(Depecker, 2002)

Elle garantit d'autre part une certaine élasticité au concept de spécialisation et permet d'envisager une analyse sous-tendant plusieurs niveaux de spécialisation et différents paramètres d'évaluation.

« Le participe passé passif présente, en effet, plusieurs intérêts, à commencer par la souplesse des interprétations : il y a place pour des degrés variables de spécialisation, de normalisation et d'intégration d'éléments exogènes [...] La notion de langue spécialisée est plus pragmatique : c'est une langue naturelle considérée en tant que vecteur de connaissances spécialisées ». (Lerat, 1995)

5. Spécialisée ou spécialisées

L'emploi du pluriel ou du singulier dépend principalement du choix du chercheur. Kocourek (1991) établit que *« faute d'arguments convaincants pour ou contre, le choix entre langue et langues exprime plutôt les préférences du chercheur ou les nuances du contexte qu'une différence fondamentale. »*.

Deux approches peuvent être mises en valeur pour tenter d'expliquer les nuances que peut engendrer l'utilisation ou pas de la désinence du pluriel. La première qui est purement théorique semble privilégier le singulier afin d'insister sur le fait que la spécialisation est un continuum et que les langues spécialisées partagent un nombre considérable de caractéristiques qu'il est possible d'étudier ; la deuxième approche, qui relève de la linguistique appliquée, quant à elle, opte pour le pluriel dans l'objectif de souligner la pluralité des langues spécialisées. Chaque langue spécialisée possède des traits spécifiques qui font son originalité et qui la différencient des autres langues spécialisées. Les deux approches sont loin d'être contradictoires, bien au contraire, elles démontrent l'acceptabilité du singulier et du pluriel dans la dénomination de la langue spécialisée et conforte l'idée de la diversité au sein de l'unité. Si chaque langue

- La langue spécialisée contrairement à la langue générale représente une langue mise en situation par un nombre de locuteurs restreint et traite d'un domaine d'activité précis. Comme nous l'avons expliqué précédemment, elle n'est pas un sous-langage avec un système de fonctionnement indépendant de la langue générale, mais simplement la langue générale en situation d'usage. Elle est par conséquent subordonnée à la langue générale et contingente parfois à la langue commune. Le degré de sa spécialisation est défini par plusieurs paramètres différents (domaine d'activité, communauté de spécialistes, etc.). Cependant, le paramètre linguistique, quant à lui, reste plus au moins commun à la majorité des langues spécialisées quelle que soit la thématique de l'énoncé, mais avec une variation due plutôt à la fréquence d'utilisation.

Même si la langue spécialisée présente quelques contraintes et limites à plusieurs niveaux, lexical, syntaxique et pragmatique, elle est loin d'être une langue contrôlée car elle continue de puiser ses ressources dans la langue générale. De plus, ces contraintes ne sont pas des contraintes au sens propre du terme pour deux raisons principales. Premièrement, les contraintes n'apparaissent pas comme typiques aux langues spécialisées, mais seulement une utilisation à une fréquence élevée des ressources proposées par la langue générale. Deuxièmement, la langue spécialisée se permet l'innovation, caractère principal des langues naturelles, et l'intégration de systèmes non linguistiques qu'elle peut traduire.

9. Langue contrôlée vs langue spécialisée

Le concept de la langue contrôlée est souvent associé à celui de la langue spécialisée, car elle présente son aboutissement et le murissement de certains de ses avantages par le biais des contraintes.

La polysémie, qui reste présente dans les langues spécialisées et traitée par la langue contrôlée, montre clairement que son rapport vis-à-vis de la langue spécialisée est un rapport d'accomplissement. En effet, la tendance du lexique spécialisé à être

monoréférentiel se retrouve améliorée par des contraintes de sémantique lexicale aboutissant à la bi-univocité des termes.

Il est important de comprendre la base de la dichotomie (langue contrôlée et langue spécialisée) pour effectuer un quelconque contrôle et concevoir le système de traduction automatique, quel type de contrôle pour quel type de texte ?

La compréhension et la reconnaissance des caractéristiques de la langue spécialisée et de la langue contrôlée permettront de mettre en place une stratégie de contrôle en concordance avec le type des énoncés et de jouer sur le niveau de contrôle en prenant en compte le degré de spécialisation des textes.

Si les deux notions partagent un héritage commun, celui de la langue naturelle, il reste à savoir quelles sont les caractéristiques distinctives de chacune.

9.1. Diachronie / Synchronie

La langue spécialisée peut être considérée à la fois sur l'axe synchronique et diachronique. Elle reste ouverte acceptant ainsi toute analyse qu'elle soit diachronique sur l'évolution de son système général et de ses règles à travers le temps ou synchronique sur le statut qu'elle tient au sein de la langue générale.

La langue contrôlée est une langue fermée et ne peut être considérée que d'un point de vue synchronique. Elle est créée une fois pour toute avec un répertoire fermé de lexique et de règles, et ne peut être modifiée que si sa conception est jugée obsolète et inadaptée. Les changements que les concepteurs apportent à la langue contrôlée se font d'une manière délibérée et demeurent contraignants car nécessitant une validation par l'ensemble de ses utilisateurs. Donc la langue contrôlée ne se situe pas sur un axe diachronique car elle est dépourvue de toute évolution sauf celle délibérément apportée.

9.2. Supériorité / Antériorité

La langue spécialisée est une langue générale en usage, donc apprise comme première langue. Elle est la langue de base que chaque personne doit apprendre afin de communiquer et d'accéder aux connaissances.

La langue contrôlée est créée par les usagers de la langue spécialisée. Elle reste donc issue de la première et se situe donc au deuxième plan. Elle doit être apprise consciencieusement par le biais de la première langue qui sert de métalangage d'accès.

9.3. Langage / Métalangage

La langue spécialisée jouit d'une flexibilité remarquable qui la distingue de la langue contrôlée. Sa capacité à produire des paraphrases lui permet de se décrire mais de décrire les autres langages. La langue des protocoles de sécurité, qui est une langue générale utilisée pour rendre compte des connaissances sur la sécurité, peut tout à fait servir à décrire d'autres domaines plus ou moins proches ou d'autres systèmes sémiotiques *« il est aussi possible d'employer la langue de spécialité en tant que métalangage, [...], c'est un système sémiotique capable de traiter les autres systèmes sémiotiques »*. (Kocourek, 1991)

La langue contrôlée ne possède pas cette capacité de paraphraser comme elle est soumise à un ensemble de contraintes avec un répertoire de lexique et de règles sensiblement limité. Elle est seulement un langage et ne peut par conséquent servir de métalangage, ni pour discourir d'elle-même ni des autres langages. La fonction de métalangage ne peut être assumée que par un langage moins contraint.

10. La langue spécialisée entre sémiotique et langue naturelle

Les spécialistes utilisent, pour exprimer les contenus spécialisés et communiquer leurs connaissances, divers systèmes sémiotiques. La langue spécialisée se voit exprimer alors à travers plusieurs codes sémiotiques qui, bien qu'ils soient opposés, conjuguent leurs ressources pour donner une cohérence à ces textes spécialisés. Les spécialistes recourent à ces signes sémiotiques pour rendre leur domaine de spécialité aisément accessible par les lecteurs mais aussi par une nécessité de combler ce que la langue naturelle à elle seule n'arrive pas à transmettre.

L'utilisation de la sémiotique se justifie par le rapport qu'entretient la langue naturelle avec la sémiotique *« la langue est un système de signes exprimant des idées, et par là comparable à l'écriture, à l'alphabet des sourds-muets, aux rites symboliques, aux formes de politesse aux signaux militaires, etc. Elle est seulement le plus important de ces systèmes. On peut donc concevoir une science qui étudie la vie des signes au sein de la vie sociale ; elle formerait une partie de la psychologie générale, nous la nommerons sémiologie (du grec semeiôn, « signe »). Elle nous apprendrait en quoi consistent les signes, quelles lois les régissent. Puisque elle n'existe pas encore, on ne peut dire ce qu'elle sera ; mais elle a le droit à l'existence, sa place est déterminée d'avance. La Linguistique n'est qu'une partie de cette science générale, les lois que découvrira la sémiologie seront applicables à la linguistique, et celle-ci se trouvera ainsi rattachée à un domaine bien défini dans l'ensemble des faits humains »*. (Saussure, 1916)

Saussure voit ainsi dans la langue une partie intégrante de la sémiologie, en interaction avec les autres codes sémiotiques. La langue spécialisée de part sa particularité de clarté et de non ambiguïté qui la démarque de la langue générale fait appel à ces moyens non linguistiques *« mais ces langues, quel que soit leur degré d'autonomie, se trouvent, au sein de la langue exposée à toutes sortes de contamination (polysémie, analogie, connotations, etc.) qui en troublent la nature et le fonctionnement. C'est pourquoi la plupart des sciences ambitionnent d'élaborer des codes non linguistiques appropriés à leur propre axiomatique »*. (Guiraud, 1971)

Chapitre 2

Protocoles de sécurité

1. La sécurité, un domaine sensible et hétéroclite

Le domaine de la sécurité est un domaine très vaste qui se situe à l'intersection de plusieurs domaines. Ceci s'explique par les enjeux majeurs, politique, économique et social qu'il représente dans les sociétés d'aujourd'hui où le risque zéro n'existe pas.

Les avancées technologiques et le progrès accomplis ces dernières décennies dans plusieurs domaines ont certainement amélioré la vie quotidienne des populations et amené un confort considérable, que ça soit dans les ménages, dans les écoles, ou les lieux de travail. Il est cependant difficile d'ignorer la face cachée de ces progrès, celle des risques et des dangers que les personnes encourent tous les jours. L'exemple le plus édifiant ces dernières années si ce n'est ces derniers mois est sans doute celui de la centrale nucléaire de Fukushima au Japon où, suite à un tsunami, un incident nucléaire a failli conduire le Japon et le reste du monde vers une catastrophe sans précédent. Dans ce genre de cas, il est important d'agir vite mais correctement, d'évaluer les risques et de donner en conséquence les consignes à suivre.

Il y a également les risques liés aux phénomènes naturels plus au moins imprévisibles : séisme, ouragan, tsunami, éruption volcanique, etc. On trouve également les risques liés à des tensions géopolitiques. Tous ces risques méritent d'être pris en compte et mesurés à temps pour pouvoir ensuite agir efficacement.

La sécurité est devenue donc, en quelques années, une préoccupation majeure des sociétés modernes et on assiste aujourd'hui à une extension fulgurante de son champ d'application. De la sécurité des lieux sensibles tels que les aéroports, à la sécurité dans les lieux de loisirs tels que les parcs, la sécurité préoccupe de plus en plus et les secteurs qui s'y intéressent ne cessent de se diversifier.

La sécurité est inhérente à toute activité humaine, plusieurs activités sont concernées par la dimension sécuritaire. Les entreprises industrielles, les collectivités locales, les hôpitaux, les compagnies aériennes, les établissements d'accueil de publics, etc., sont autant de structures qui se préoccupent de la sécurité des personnes et des

biens. Il est légitime aujourd'hui de constater que l'objectif principal est de ne plus laisser de place à l'insécurité.

En effet, l'étendue des risques aussi diversifiés que les secteurs auxquels ils sont liés témoigne du caractère inter-domanial de la sécurité et prouve encore une fois la nécessité de se poser certaines questions quant aux moyens de communication à mettre en place pour la gestion des crises.

La communication dans le domaine de la sécurité se fait de plusieurs façons et par plusieurs moyens, les messages peuvent être transmis sous forme d'alertes ou à travers des protocoles de sécurité et sont rédigés par des spécialistes du domaine de la sécurité. Ils peuvent être sonores ou visuels, et comportent des caractéristiques linguistiques spécifiques. Un bon message est un message simple, clair et efficace, il doit répondre aux besoins et aux attentes des récepteurs et être compris par tous. Par conséquent, la rédaction joue un rôle majeur dans la portée communicative des messages, et reste un facteur déterminant dans la compréhension du message.

2. Corpus de travail

Le corpus sur lequel nous nous sommes appuyés provient de différents partenaires avec qui nous avons collaboré dans le cadre du projet LiSe. La sélection de ce corpus s'est faite par rapport aux objectifs que nous nous sommes fixés à savoir, la création d'une langue contrôlée dans le domaine de la sécurité et la construction d'un système de traduction automatique français-arabe basé sur un modèle linguistique isomorphique.

Le choix est donc guidé par les applications que nous entendons concevoir, ce qui nous conduit à la définition de Rastier (2001) du corpus « *Un corpus est un regroupement structuré de textes intégraux, documentés, éventuellement enrichis par des étiquetages, et rassemblés : de manière théorique réflexive en tenant compte des discours et des genres, et de manière pratique en vue d'une gamme d'applications.* »

La définition de Rastier nous semble pertinente car convenant à notre méthodologie pour plusieurs raisons :

- Elle met au cœur des analyses les productions effectives de la langue qui ne sont pas que des mots, des phrases mais aussi des textes. Les mots et les phrases ne peuvent être isolés et même quand ils ont été isolés, ils restent rattachés à un genre ou à un discours. Les mots et les phrases que nous analysons dans notre corpus, même isolés dans les cas des messages d'alerte et des protocoles de sécurité, appartiennent au même genre en l'occurrence les textes de sécurité.
- Elle concilie les différentes approches issues de la linguistique du corpus. Elle intègre aussi bien les études sur le lexique⁹ qui prétendent à l'exhaustivité (corpus lexicographique ou sacs de mots) que celles sur les unités grammaticales (phrases)¹⁰ puisque le texte comme unité minimale du corpus inclut les phrases et les mots. Le contrôle et la conception du modèle linguistique qui permettra le passage du français vers l'arabe imposent la prise en compte de toutes les unités de la langue, les mots, les phrases ainsi que les textes.
- Le choix des textes dépend des applications auxquelles nous voulons arriver. Cette dépendance vis-à-vis d'une application ou d'un ensemble d'applications apaise la polémique autour de la notion de représentativité du corpus. En effet aucun corpus ne peut prétendre représenter la langue entière, toutefois un corpus convient ou pas à une tâche précise et c'est seulement en lien avec cette tâche qu'on peut juger de sa représentativité et de son homogénéité. Dans notre cas, le choix du corpus était guidé par les objectifs que nous nous sommes fixés : créer une langue contrôlée, et construire un système de traduction automatique du français vers l'arabe, appliqué au domaine de la sécurité. Le travail sur le corpus était par conséquent un travail en intension et non en extension. (Voir chap1/7.3. Spécialisation par traits linguistiques)

Le choix de ce corpus se devait donc de respecter des critères de représentativité et d'homogénéité qui permettaient un travail d'analyse précis. Ces critères ont été atteints

⁹ J.M. Sinclair, *Corpus Concordance Collocation*, Oxford, Oxford University Press, 1991.

¹⁰ S. Hunston and G. Francis, *Pattern Grammar. A Corpus-driven Approach to the Lexical Grammar of English*, Amsterdam and Philadelphia, Benjamins, 2000.

grâce d'une part à nos différents partenaires et aux différents domaines qu'ils représentaient (aéronautique, médecine et sécurité civile) mais qui appartiennent au même genre (sécurité). La sécurité comme nous l'avons souligné précédemment concerne de nombreux domaines, la diversité des partenaires nous a d'ailleurs permis d'en avoir l'accès et de traiter plusieurs textes. D'autre part, la constitution du corpus a été précédée par une identification des objectifs recherchés.

Le corpus sur lequel nous nous appuyons est une sorte d'observatoire où les faits empiriques du langage utilisé dans la sécurité sont traduits en un modèle théorique sous tendant la langue contrôlée et le système de TA français-arabe.

Ce passage de pratique discursive à un modèle théorique passe par :

- la compréhension du fonctionnement interne de chaque texte,
- la définition des caractéristiques linguistiques communes,
- le repérage des récurrences et des régularités de certains phénomènes linguistiques,
- la mesure de l'étendue des ambiguïtés présentes dans les textes,
- la mise en place un dispositif de contrôle *statique* et *dynamique*,
- la construction à partir de la langue contrôlée d'un modèle syntaxico-sémantique abstrait.

Le corpus que nous avons collecté se divise en deux grandes catégories, les protocoles de sécurité et les messages d'alerte. Les protocoles de sécurité touchent trois domaines différents (aéronautique, médecine et sécurité civile) est sont répartis en deux catégories distinctes selon le public auquel ils sont destinés (grand public ou professionnels).

2.1. Protocoles

Plus de 400 protocoles ont été collectés, ils proviennent de différentes sources, aussi bien des sites officiels (ministères par exemple), des organismes internationaux (ONU, OMS, ...) ou directement de nos partenaires.

Les protocoles de sécurité sur lesquels nous avons travaillé concernent divers domaines (voir chap2/ 3. Domaines de travail) et sont tour à tour liés au grand public ou à un public professionnel. Voici ci-dessous la liste des protocoles sur lesquels nous avons travaillé classés par domaines et par publics concernés.

Sécurité civile

Public visé : grand public

- Catastrophes naturelles : ouragan, séisme, éruption volcanique, tsunami, etc.
- Catastrophes industrielles : risques chimiques, risques biologiques, etc.
- Accidents domestiques : incendie, brûlure, blessure, etc.
- Sécurité routière
- Sécurité au travail
- Attaques : terrorisme, animaux, etc.

Public visé : professionnels

- Catastrophes naturelles : ouragan, séisme, éruption volcanique, tsunami, etc.
- Catastrophes industrielles : risques chimiques, risques biologiques, etc.
- Accidents domestiques : incendie, brûlures, chute, etc.
- Secourisme
- Sécurité routière
- Attaques : terrorisme

Santé

Public visé : grand public

- Prévention : grippe A, MST, etc.
- Vaccination
- Régimes alimentaires

Public visé : professionnels

- Procédures médicales
- Gestion de crises

Aéronautique

Public visé : grand public

- Sécurité des passagers dans les avions

Public visé : professionnels

- Consignes de sécurité aux pilotes : atterrissage, décollage, etc.

2.2. Messages d'alerte

Les messages d'alertes sont généralement des messages d'action capitale. Ils ne sont émis que dans des conditions exceptionnelles, ce qui explique leur rareté par rapport aux protocoles de sécurité. Les messages d'alerte sont destinés au grand public et sont dans la majorité des cas des messages courts (écrits, oraux, visuels) et d'une importance cruciale, le degré d'urgence étant élevé. Les alertes sont souvent diffusées à temps réel comme c'est le cas pour les PMV (les Panneaux à Messages Variables) ou les alertes enlèvements.

3.1. Sécurité civile

La sécurité civile représente l'ensemble des moyens mis par l'état pour prévenir et protéger l'environnement, les citoyens et les biens contre différents risques (catastrophes naturelles, catastrophes industrielles, épidémie, terrorisme, etc.). La sécurité civile, c'est aussi la formation des populations aux gestes de premiers secours dans la perspective de replacer le citoyen au cœur du dispositif de la sécurité civile. Nous nous référons pour cette définition à la loi promulguée le 13 août 2004 visant la modernisation de la sécurité civile.

« La sécurité civile a pour objet la prévention des risques de toute nature, l'information et l'alerte des populations ainsi que la protection des personnes, des biens et de l'environnement contre les accidents, les sinistres et les catastrophes par la préparation et la mise en œuvre de mesures et de moyens appropriés relevant de l'Etat, des collectivités territoriales et des autres personnes publiques ou privées. Elle concourt à la protection générale des populations, en lien avec la sécurité intérieure au sens de la [loi n° 2003-239 du 18 mars 2003](#) pour la sécurité intérieure et avec la défense civile dans les conditions prévues par [l'ordonnance n° 59-147 du 7 janvier 1959](#) portant organisation générale de la défense [...] »

Le domaine de la sécurité civile englobe plusieurs sous-domaines et concerne aussi bien les professionnels *« Les missions de sécurité civile sont assurées principalement par les sapeurs-pompiers professionnels et volontaires des services d'incendie et de secours ainsi que par les personnels des services de l'Etat et les militaires des unités qui en sont investis à titre permanent. Concourent également à l'accomplissement des missions de la sécurité civile les militaires des armées et de la gendarmerie nationale, les personnels de la police nationale et les agents de l'Etat, des collectivités territoriales et des établissements et organismes publics ou privés appelés à exercer des missions se rapportant à la protection des populations ou au maintien de la continuité de la vie nationale, les membres des associations ayant la sécurité civile dans leur objet social ainsi que les réservistes de la sécurité civile. »* (Article 2) que le grand public *« Toute personne concourt par son comportement à la sécurité civile. En fonction des situations auxquelles elle est confrontée*

et dans la mesure de ses possibilités, elle veille à prévenir les services de secours et à prendre les premières dispositions nécessaires. » (Article 4)

De nos jours, les sources de risques sont multiples. Le risque zéro n'existe pas et chacun de nous peut être confronté à une situation de risque immédiat ou imminent. Que ce soit à domicile, sur les lieux de travail ou dans des lieux publics, le phénomène de l'insécurité gagne du terrain et les risques sont de plus en plus présents dans nos sociétés modernes (terrorisme, violence urbaines).

La sécurité civile étant au carrefour de plusieurs domaines et étant la préoccupation du professionnel comme du simple citoyen, la collecte du corpus était donc aisée et le nombre de protocoles récupérés est supérieur à celui des autres domaines. Les protocoles que nous traitons dans notre travail, sont aussi variés que les domaines rattachés à la sécurité civile : incendie, terrorisme, catastrophe nucléaire, grippe H1N1, séisme, etc.

3.2. Médecine

Nul ne peut nier le lien étroit entre médecine et société : la médecine est aujourd'hui, plus sociale qu'individuelle. La profession médicale ne se résume plus aux simples relations médecin/patient, elle est également en interaction permanente avec la société dans laquelle elle prend ses repères et se développe. En effet, la médecine a pris aujourd'hui une dimension sociale et humaine incontournable. L'exercice professionnel de la médecine que ce soit en termes d'approfondissement des connaissances ou dans la manière d'exercer la profession a connu et connaît encore des mutations profondes dues aux évolutions de la société et à l'émergence des nouvelles méthodes et technologies.

En effet, l'exercice quotidien de la médecine est confronté à des changements constants et à des modalités de pratiques très variées. Le champ professionnel s'élargit de plus en plus alliant réseau social et réseau sanitaire avec l'apparition de nouvelles relations comme celles établies avec les professions paramédicales.

En outre, être médecin aujourd'hui n'est pas seulement posséder des compétences médicales mais c'est aussi être pétri de connaissances médico-sociales et juridiques. Les mutations qu'a subies la médecine ont, en effet, eu des répercussions sur le cadre réglementaire de la profession. Le contexte des relations entre médecin et patient est caractérisé par un cadre juridique rigide et la médecine d'aujourd'hui se voit encadrée par une batterie de réglementations, des décrets, des arrêtés, et des lois de tout genre, veillant à ce que chaque acte médical soit conforme aux procédures recommandées dans l'exercice médical et garantissent sa qualité. Le patient attend de la part de son médecin non seulement un dévouement et des compétences mais surtout des actes médicaux sûrs et sans risques dans un climat de confiance. En d'autres termes, il s'agit d'éviter des erreurs médicales et de définir les responsabilités de chacun.

Les erreurs médicales peuvent, entre autres, provenir d'une mauvaise interprétation d'une procédure décrite dans un protocole médical. La personne, médecin ou professionnel paramédical, en charge d'effectuer l'acte médical peut être confrontée à une incompréhension face à certaines instructions ou descriptions écrites. Elle peut aussi mal interpréter des indications parce qu'elles sont ambiguës, pouvant provoquer des accidents médicaux lourds de conséquences, d'où l'intérêt de contrôler et d'uniformiser les procédures médicales. Les rédactions doivent être claires, précises et sans ambiguïtés pour une meilleure exécution de l'acte médical. Un éventuel transfert du savoir médical pourrait également être simplifié par l'utilisation de cette rédaction normalisée car la traduction des protocoles médicaux devient plus fiable. Nous avons, au cours de notre travail, traité des problèmes liés à la traduction de ces protocoles qui ne diffèrent pas d'ailleurs de ceux des protocoles de sécurité civile ou de ceux de l'aéronautique. L'objectif est d'avoir une vue d'ensemble des textes de sécurité et d'établir un modèle de TA français-arabe.

3.3. Aéronautique

L'aéronautique ne diffère pas des autres domaines sensibles, nous irons même à le considérer comme l'un des domaines nécessitant plus de sécurité étant donné la

multiplicité des paramètres à prendre en compte : la sécurité des passagers, du personnel, du matériel, la sécurité au sol, pendant le vol, etc. Le domaine de l'aéronautique est également le domaine par excellence où la réalisation des différentes tâches dépend entièrement de la communication entre les experts du domaine. Le nombre des langues contrôlées établies dans ce domaine témoigne de la nécessité d'éviter toute ambiguïté du langage naturel et l'importance de réguler les communications. Que ce soit pour le personnel naviguant, pour les contrôleurs aériens ou les pilotes, l'urgence était de mettre en place des langages compréhensibles et identiques pour tous, afin d'éviter de potentiels drames aériens.

En effet, l'utilisation du langage naturel en milieu professionnel génère des tensions et comporte des risques qui peuvent s'avérer fatals d'autant plus que les risques liés aux activités professionnelles sont aussi élevés que ceux que comporte l'exercice même de ces différentes activités *« l'usage du langage et des langues naturelles au travail tend à produire des approximations, des malentendus ou des incompréhensions entre professionnels engagés dans des tâches communes. C'est en ce sens qu'on peut parler d'un "risque du langage" au travail »* (Vergely, 2008).

La langue de référence dans le domaine de l'aéronautique est l'anglais, car l'objectif des organisations internationales est d'uniformiser le langage de l'aéronautique. De ce fait, les protocoles que nous avons pu nous procurer sont en anglais et sont déjà contrôlés. Au premier abord, cela semble donc peu intéressant, or nous nous sommes rendu compte que l'exploitation de ces protocoles n'est pas sans intérêt pour nous, surtout pour ce qui est de la traduction automatique et de la création du modèle linguistique.

Certaines règles de contrôle mises en place sur l'anglais s'avèrent souvent applicables à la langue française, le contrôle est donc plus au moins universel. Il était aussi possible pour nous, à travers les textes contrôlés, de se projeter dans le contrôle du français et d'imaginer un modèle qui pourrait concilier les deux langues, le français et l'arabe.

4. La communication écrite en milieu professionnel, un facteur de risque ?

Les écrits de travail jouent de plus en plus un rôle capital et ils s'avèrent être l'outil indispensable pour le transfert des connaissances nécessaires à l'accomplissement des tâches et le déroulement des diverses actions, surtout quand ces dernières revêtent une fonction de procédure ou d'information.

Cependant, et contrairement à une idée bien ancrée qui tend à associer le risque professionnel uniquement à l'usage de l'oral, l'usage de la langue naturelle écrite en situation de travail peut comporter elle aussi des risques. Ces risques sont d'ailleurs bien plus importants que ceux de la communication orale car le rédacteur et le récepteur ne sont pas forcément dans une relation directe qui pourrait dissiper d'éventuels malentendus.

Les risques professionnels ne sont donc pas l'apanage de l'oral, l'écrit aussi peut générer des situations de risques dues à une mauvaise communication entre les différents experts du domaine.

Il est clair cependant que certains domaines sont plus sensibles que d'autres, parce qu'ils exigent d'une part une visée communicative non ambiguë, et d'autre part parce que ces domaines sont à la pointe des technologies comme par exemple les domaines de l'aéronautique, du nucléaire ou de la médecine, jugés « domaines tendus », la tension liée à l'usage du langage naturel étant constamment élevée.

De surcroît le langage naturel, qu'utilisent les professionnels de ces domaines supposé être un langage spécialisé, n'échappe pas aux différents problèmes de la langue générale, plus particulièrement ceux liés à son fonctionnement sémantique (polysémies, ambiguïtés, etc.). La langue spécialisée, comme nous l'avons développé précédemment, comporte dans ses échanges langagiers un nombre de phénomènes typiques au langage naturel qui crée un décalage entre l'information transmise, pensée comme information

claire, stable et transparente par son rédacteur et la réalité langagière qui fait intervenir le contexte et la variation.

Les risques liés au langage naturel sont d'ores et déjà identifiés, comme le souligne Vergely (2008), par les organisations qui ont réfléchi aux moyens qui limiteraient ces risques comme par exemple la mise en place de la phraséologie aéronautique¹³ dans les communications entre pilotes et contrôleurs de la navigation aérienne. En effet, plusieurs solutions sont en mesure d'endiguer les risques que peut induire l'usage du langage naturel ainsi que la transmission des connaissances dans d'autres langues par la traduction automatique. L'une de ces solutions consiste à créer une langue contrôlée appliquée à la langue spécialisée qui permet des communications sûres, claires et d'une efficacité optimale entre rédacteurs et utilisateurs finaux.

C'est en effet, dans cette perspective que nous situons notre travail, chercher dans un domaine circonscrit, le domaine de sécurité en l'occurrence, des régularités et des normes langagières. Ces dernières seraient suffisamment récurrentes tant au niveau du contenu qu'au niveau de la forme pour permettre la mise en place d'un modèle linguistique formel et contrôlé. Ce modèle identifie les problèmes monolingues liés à la compréhension et les problèmes contrastifs émanant de la mise en contact des deux langues, le français et l'arabe. Il permettra la rédaction de protocoles de sécurité non ambigus ainsi que leur traduction vers l'arabe.

5. Protocoles de sécurité, des textes procéduraux ?

Nous nous intéresserons, dans cette partie, à la structure interne et aux régularités linguistiques générales des protocoles de sécurité identifiés comme procéduraux, et aux domaines¹⁴ de la sécurité en général, transposables à une variété de textes. Cette analyse est réalisée sans prétendre entrer dans les détails de chaque domaine ou faire une

¹³ Le terme phraséologie ne renvoie pas au sens habituellement utilisé en linguistique mais au langage contrôlé instauré pour faciliter les communications radio-téléphoniques entre pilotes et contrôleurs de la navigation aérienne.

¹⁴ Médecine, aéronautique, sécurité civile

présence d'invariants structurels confirment ce lien étroit. Ils permettent, par conséquent, une modélisation de leur fonctionnement, dont dépendront la conception de la langue contrôlée et la traduction automatique français-arabe.

7. Définition des protocoles de sécurité

Après s'être attardé sur la classe des textes procéduraux et montré le flottement de terminologie dont souffre cette catégorie de textes, nous considérons les protocoles de sécurité comme textes à dominance procédurale.

Nous les définissons comme suit : un protocole de sécurité est un texte procédural qui a pour principal objectif la prescription d'un ensemble d'actions destiné au grand public ou aux professionnels selon qu'on vise la vulgarisation ou non, afin de garantir un climat de sécurité immédiat ou prévisionnel au vu de certaines conditions.

Nous comprenons alors l'importance de ces textes dans la vie quotidienne et l'intérêt que cela représente pour la communauté professionnelle. En s'assurant que l'environnement est propice et que les conditions sont réunies, les protocoles de sécurité contribuent au bon déroulement des actions et facilitent la réalisation des différentes tâches que le sujet doit accomplir. Bien au delà des enjeux sécuritaires, des enjeux légaux peuvent faire surface dans ce type de texte d'où l'importance de réduire à zéro le taux d'erreurs.

8. Structure externe des protocoles de sécurité

La structure externe des protocoles de sécurité est généralement stable, elle est constituée d'un ensemble cohérent de conditions-actions, et son organisation textuelle est régie par des éléments linguistiques récurrents. Le caractère structuré des textes de sécurité apporte une pérennité à la communication, le rédacteur des protocoles de

9.3.5. Approche analytique et approche synthétique

Les constructions à verbes support peuvent être appréhendées de deux manières différentes :

- L'approche synthétique suppose que le verbe support est une unité composée, et c'est en tant que telle que sa valence doit être décrite. Le verbe *prendre feu* sera décrit comme verbe intransitif alors que le verbe *avoir mal à* sera décrit comme un verbe transitif indirect. Aucun statut syntaxique n'est attribué dans ce cas aux noms *feu* et *mal*
- L'approche analytique part d'une analyse syntaxique interne et assigne un statut syntaxique aux noms prédicatifs. Elle met en parallèle le second élément de la construction à support et celui du nom objet des verbes distributionnels.

Or, les deux approches s'enchevêtrent au moment de la traduction du français vers l'arabe, il y a un jeu d'équivalence entre constructions analytiques et constructions synthétiques. C'est pour cela d'ailleurs qu'il est impératif à notre sens de considérer que les deux approches ne s'opposent pas l'une à l'autre, mais se complètent. Nous estimons par ailleurs que le blocage et l'ouverture des substitutions paradigmatiques, d'un point de vue intra- et interlangue, est l'approche convenable à adopter. Dans une perspective de traduction automatique vers l'arabe, l'approche synthétique conviendra mieux à des expressions telles que : *prendre soin* (اعتناء), *prendre feu* (احتراق) car les équivalences sont monolexicales en arabe, en revanche pour des expressions telles que : *prendre l'escalier* (استخدام السلم), *avoir froid* (شعور بالبرد), l'approche analytique sera la mieux adaptée car les équivalences en langue arabe restent polylexicales et les substitutions paradigmatiques n'affectent pas toujours le verbe support.

9.3.6. Verbes support & verbes étendus

La liberté paradigmatique relative qu'autorisent les verbes support reste en corrélation directe avec la liberté d'agencement au sein de la phrase. Sachant que les

Chapitre 3

LANGUE CONTROLEE EN VUE DE LA TRADUCTION AUTOMATIQUE VERS L'ARABE

I. Langues contrôlées bref aperçu

1. Langues contrôlées, des besoins croissants

L'industrie, dont les activités s'étendent de l'extraction des matières premières à la fabrication de produits finis, est génératrice d'une quantité importante de documentation aussi variée que les différents domaines auxquels elle est destinée. La médecine, dont les progrès et la recherche sont de plus en plus mondialisés, ainsi que le domaine de la sécurité civile, qui s'adresse de plus en plus à des populations qui se déplacent et qui voyagent aux quatre coins du globe, génèrent également tous deux un grand nombre de documents d'information et de notices diverses.

La gestion de cette documentation doit répondre à bon nombre d'exigences. En effet, celle-ci doit être accessible, régulièrement tenue à jour et facile à utiliser. Elle est de plus souvent multilingue, cela constitue un défi permanent notamment dans le domaine de la traduction automatique.

Dès le début du XX^e siècle, l'idée d'une simplification et d'une standardisation, afin de rendre toute cette source d'inspiration plus lisible et plus facilement exploitable, a fait son apparition. Les langues contrôlées ont peu à peu fait leur apparition et font désormais partie intégrante de la panoplie d'outils documentaires conçus pour faciliter la gestion de ce flux de données parfois très techniques. Elles se sont d'ailleurs imposées comme un outil indispensable. En effet, un document technique (mode opératoire, rapport technique, bulletin de service, etc.) doit être : clair et précis (pas d'interprétation), concis, neutre, objectif et explicite, apte au traitement informatique (notamment pour la traduction automatique), facile à mettre à jour, conforme aux normes et apte à faire autorité.

La plupart des langues contrôlées proviennent directement d'entreprises industrielles privées où elles sont développées « en interne » pour répondre aux besoins « internes ». Aujourd'hui le monde professionnel utilise massivement et de manière quotidienne les langues contrôlées, elles permettent en effet de répondre aux besoins de communication parfois très technique au sein de l'entreprise mais aussi avec les différents partenaires de cette dernière. La plupart du temps, ces langues contrôlées sont développées non pas par des linguistes mais par des techniciens et par des

documentalistes industriels. Ce sont des langues spécialisées régies par diverses règles plus ou moins formalisées complétant les règles standards. Ces restrictions portent sur le lexique, la syntaxe ou encore sur l'agencement et la mise en forme des textes. Malheureusement, et du fait de leur développement « en interne », les langues contrôlées restent très confidentielles et très diverses quant aux règles et aux restrictions qu'elles imposent. Il n'existe par conséquent pas une seule langue contrôlée mais bien un ensemble de plusieurs, toutes liées à une entreprise ou à un domaine. Certaines sont d'usage purement interne, d'autres sont devenues des spécifications interprofessionnelles.

2. Quelques langues contrôlées existantes

Nous allons voir à présent, d'une manière succincte, quelques-unes de ces langues contrôlées :

- AECMA Simplified English

Exemple le plus connu des langues contrôlées, elle a été conçue dans les années 80 dans le but de rendre accessibles les documents de maintenance aéronautique à des locuteurs non anglophones (niveau d'anglais limité). C'est la langue de référence pour la rédaction des documents aéronautiques et pour les projets commerciaux et militaires internationaux.

- Français Rationalisé (GIFAS)

En France, la documentation technique doit être fournie dans sa version française en plus de sa version originale (si elle est écrite en langue étrangère). Le Français Rationalisé a été conçu à la fin des années 80 (suite à la mise en place de l'anglais simplifié) par le Groupement des Industries Françaises Aéronautiques et Spatiales (GIFAS). Elle permet de traduire plus facilement et d'améliorer la compréhensibilité.

II. Pourquoi contrôler ?

1. Définition

La langue contrôlée est directement liée au domaine d'utilisation pour lequel elle a été créée. La langue contrôlée est un ensemble de règles qui permet de réduire les ambiguïtés. Elle est en fait un sous-ensemble d'une langue naturelle, dont le lexique, la grammaire et la sémantique ont été volontairement restreints. Elle se présente le plus souvent sous la forme d'un guide rédactionnel et d'un dictionnaire de termes préconisés.

La langue contrôlée cherche à remplir plusieurs objectifs qui ont tous pour but une désambiguïsation de la langue naturelle en vue de l'amélioration de la compréhension de textes techniques et en vue de favoriser sa traduction humaine comme automatique.

La langue contrôlée permet d'assurer une meilleure diffusion de l'information. Plus précise, plus rapide et efficace, elle assure une bonne compréhension du texte qu'elle a formalisée en prévenant notamment les erreurs d'interprétation dans le cadre de situations d'urgence ou lors d'états de stress. Elle permet des traductions correctes grâce à la suppression d'un maximum d'ambiguïtés, et permet d'obtenir des textes uniformisés, conformes à des modèles et des normes établis avec des professionnels.

Les caractéristiques principales des langues contrôlées sont l'élimination des ambiguïtés de quelques origines qu'elles soient, l'utilisation d'un lexique restreint permettant d'éviter les ambiguïtés lexicales (pour une lexie, on a un seul sens et une seule catégorie grammaticale), l'élaboration de structures syntaxiques restreintes et certaines règles de mises en forme des textes traités par la langue contrôlée.

Nous traiterons dans la suite de ce chapitre les différents types d'ambiguïtés rencontrées et les solutions que nous avons choisies afin de limiter ces problèmes et d'offrir un travail de traduction de bonne qualité.

Les structures syntaxiques restreintes, quant à elles, feront l'objet d'un chapitre particulier (Voir chap7/6. Groupes verbaux du français contrôlé).

III. Langue contrôlée et analyse comparative

1. Introduction

La construction d'un système de traduction automatique se heurte à une série de difficultés qui peuvent être rangées en deux catégories principales.

- Les premières sont monolingues, statiques, elles proviennent de la difficulté de modéliser avec précision les connaissances formelles sur la langue source et cible. En d'autres termes, il est difficile de créer un schéma biunivoque entre représentations sémantiques et représentations syntaxiques, la langue étant en effet un ensemble de relations rarement bijectives entre formes syntaxiques et concepts sémantiques.
- Les secondes sont contrastives, dynamiques et concernent la mise en contact des deux langues et le lien qu'il faut établir pour effectuer la traduction automatique (Voir chap6/7. Difficultés de traduire automatiquement).

Il est donc nécessaire de disposer d'un modèle syntaxique et d'un modèle sémantique bien définis capables d'assumer le passage de la langue source à la langue cible. Le modèle syntaxique est chargé de représenter les structures syntaxiques abstraites de la langue et de recenser toutes les constructions que le système de traduction prendra en charge. La reconnaissance parfaite et efficace de la syntaxe d'une langue passe par la résolution des écueils de l'explosion combinatoire qui génère des ambiguïtés structurales et cause un temps d'analyse insoutenable. Le modèle sémantique doit être souple, permettant une représentation fine de la sémantique de chaque langue, et doit être autant que le modèle syntaxique en parfaite harmonie avec la nature linguistique des langues en question. Autrement dit, la formalisation linguistique doit respecter les aspects langagiers de chaque langue : langue de départ et langue d'arrivée.

Une fois les deux modèles établis, il est essentiel de définir une théorie qui fera le lien entre les deux modèles, syntaxique et sémantique, qui sont indépendants l'un de l'autre au départ. Cette partie de modélisation est très délicate de part le rôle décisif qu'elle joue. C'est elle qui déterminera la finesse du système de traduction automatique et sa qualité en termes de précision, de rapidité et d'efficacité.

L'architecture du système TACCT, nous le montrerons d'ailleurs à travers ce sous-chapitre, puise sa force de deux principales orientations : langage contrôlé et analyse comparative.

Ces deux orientations sont complémentaires et interagissent au point qu'il est difficile de poser des frontières entre les deux. Le modèle syntaxico-sémantique adopté pour la traduction est le fruit de cette nouvelle approche bipolaire. Celle-ci est basée d'un côté sur l'étude des divergences entre le français et l'arabe menée sur un large corpus d'alertes et de protocoles de sécurité bilingues et d'un autre côté sur le contrôle. Les deux modèles syntaxique et sémantique de chaque langue sont construits simultanément non seulement en prenant en compte le particularisme de chacune d'elles, mais aussi en s'intéressant aux spécificités émanant de leur contact.

2. Langue contrôlée et TA

2.1. Langue contrôlée LiSe et système TACCT

La langue contrôlée LiSe dans le système TACCT est créée par le système et pour le système lui-même. Sa création répond exactement aux besoins de traduction automatique français-arabe et aux objectifs fixés pendant la conception du système à savoir l'efficacité, la précision et la rapidité. La langue contrôlée LiSe vise à augmenter la lisibilité des protocoles de sécurité et à éviter au système de traduction automatique une panoplie d'ambiguïtés dues principalement à la complexité du langage humain. Elle tend à simplifier à la fois la langue de départ et la langue d'arrivée, tout en préservant leurs richesses et leurs particularités. Elle vise aussi à donner au rédacteur un outil normatif de rédaction, afin que les textes soient standardisés et conformes aux normes de lisibilité et de non ambiguïté.

Langue contrôlée LiSe et système TACCT sont indissociables. Le contenu de la langue contrôlée LiSe est le reflet de l'architecture du système TACCT français-arabe, il correspond au travail d'analyse et de réflexion mené lors de sa conception. Inversement,

Chapitre 4

SYSTEME LINGUISTIQUE

ARABE

1. Introduction

Nous ne pouvons concevoir un traitement automatique de la langue naturelle sans connaître son fonctionnement interne et sans maîtriser ses manifestations linguistiques. Pour les besoins des langues contrôlées orientées humain et machine et la traduction automatique, nous estimons que la maîtrise de la langue passe par une analyse microsystemique à la fois monolingue et contrastive de l'ensemble de la langue. La langue est un système constitué de microsystemes en interaction permanente qui ne peut être analysée que si nous procédons à sa description, microsysteme par microsysteme.

Ce chapitre propose de décrire le mode de fonctionnement de la langue arabe en se basant sur une approche microsystemique fragmentaire. Il ne prétend néanmoins pas être une étude complète de langue arabe, mais un récapitulatif des principaux éléments moteurs de cette langue. Nous considérons qu'un travail en intensification est l'atout principal de notre recherche. Il définit avec précision, en rapport avec les besoins de la langue contrôlée et de la traduction automatique contrôlée français-arabe, les problèmes posés par la langue ainsi que leurs solutions, sans recourir à des études holistiques, source de bruits et d'inexactitudes. Chaque microsysteme sera donc décrit et mis en relation avec d'autres microsystemes pour pouvoir saisir les différentes interactions. Ainsi nous décrirons par exemple les substantifs en relation avec les adjectifs et les cas, les constructions nominales et verbales en relation avec les verbes et les noms, etc.

Les descriptions seront souvent accompagnées d'exemples tirés du corpus que nous avons traité, hormis quelques-uns que nous avons identifiés par deux astérisques (**). Tous les exemples sont traduits en français car l'analyse prend comme repère le français.

En effet, la description de la langue arabe doit se faire en vue de la traduction automatique français-arabe et dans l'optique de mettre en exergue les divergences ainsi que les convergences entre ces deux langues. Ainsi, il sera possible de définir quelles règles de contrôle doivent être créées pour aplanir les écarts et quelles règles de

transfert appliquer pour arriver à un modèle syntaxique et sémantique isomorphe qui permet des traductions automatiques fiables et rapides.

2. La phonétique dans la langue arabe

2.1. L'alphabet et caractères spéciaux

Il existe en arabe 28 lettres. Ces lettres s'écrivent de droite à gauche. Lorsqu'elles sont liées et selon leurs positions dans le mot (au début, initiales ; au milieu, médianes ; à la fin, finales ou quand elles sont isolées), elles changent légèrement de forme. Certaines lettres ne se lient jamais à celle qui suit (و - ز - ر - ذ - د - لا). Elles n'existent que sous deux formes : isolées et finales.

Il existe deux types de voyelles :

- les brèves. Elles peuvent disparaître dans l'écriture mais apparaissent dans la lecture. Ce sont :
 - la damma ء,
 - la fatha ا
 - la kasra ـ
- les longues. Pour avoir des voyelles longues qui durent deux fois plus que les voyelles brèves, on fait suivre les voyelles brèves de lettres de prolongation.
 - la damma d'un و : ءو
 - la fatha d'un ا ou d'un ي : اء / يء
 - la kasra d'un ي : يـ

Les lettres de prolongation sont le :

- ا : Il n'est jamais consonne, il sert à marquer une voyelle longue (ء / اء), sauf s'il est le support d'une hamza ou la marque d'un tanwin.

- و et ي : elles peuvent être des lettres de prolongation marquant des voyelles longues و et ي ou des consonnes waw (و) et ya (ي) d'où la dénomination semi-consonnes. (Voir le tableau ci-dessous)

D'autres signes peuvent apparaître dans le mot, le sukûn ْ qui correspond à un repos indique que la lettre est sans voyelle. Une consonne quelconque, comme en français, peut être renforcée. Ce renforcement est appelé gémination et est indiqué par la sādda ّ.

On trouve deux sortes de syllabes en arabe :

- les brèves, formées d'une consonne ou d'une semi-consonne suivie d'une voyelle. سَقَطَ (trois syllabes brèves).
- les longues, formées d'une consonne ou d'une semi-consonne avec une voyelle brève suivie d'une consonne instable quiescente. نَسْفُطُ (une syllabe longue + 2 syllabes brèves).

L'écriture de la hamza (ء) obéit à des règles précises qui dépendent de la place qu'elle occupe dans le mot. Elle est aussi en rapport étroit avec l'émission d'une voyelle. On l'écrit presque toujours comme une voyelle longue que l'on appelle support du hamza (*alif* (ا) : أ ou إ, waw (و) : و, ya (ي) sans points: ى). Elle peut être écrite sans support aussi (هـ). Deux hamza ne peuvent pas être écrites à la suite l'une de l'autre, ni une hamza suivie d'un alif. On se sert alors d'une madda (prolongation) qui s'écrit au-dessus d'un *alif* ٱ.

En arabe, la notion de racine a une très grande importance comme dans toutes les langues sémitiques. Une racine est une suite de consonnes présentées dans un ordre fixe. La plupart des racines sont composées de trois lettres (trilitères). Il existe aussi quelques racines de deux lettres (bilitères) ou de quatre lettres (quadrilitères). Elles expriment une idée générale. La notion de schème est aussi importante que celle de la racine. Une racine peut être manipulée sous différentes formes à l'aide de voyelles, d'ajout de lettres, etc. afin de former des mots selon des modèles précis que l'on appelle schèmes.

2.2. La prononciation et l'écriture de chaque lettre de l'alphabet

Les lettres arabes s'écrivent différemment selon qu'elles soient :

1. finales
2. médianes
3. initiales
4. isolées

Formes				Noms	Trans- cription	Valeur en français et prononciation
1	2	3	4			
			ء	Hamza	‘	attaque vocalique forte (<i>assez !</i>) ou explosive glottale
ا	ا	ا	ا	'alif	â	â, voyelle longue
ب	ب	ب	ب	bâ'	b	b, labiale occlusive sonore
ت	ت	ت	ت	tâ'	t	t, dentale occlusive sourde
ث	ث	ث	ث	ṭâ'	ṭ	interdentale, spirante sourde
ج	ج	ج	ج	jîm	j	j, spirante palatale sonore (<i>dj</i> en Algérie et Proche-Orient, <i>g</i> en Egypte)
ح	ح	ح	ح	ḥâ'	ḥ	spirante laryngale sourde
خ	خ	خ	خ	ḫâ'	ḫ	vélaire spirante sourde
د	د	د	د	dâl	d	d, dentale occlusive sonore
ذ	ذ	ذ	ذ	ḏâl	ḏ	interdentale spirante sonore
ر	ر	ر	ر	râ'	r	r « roulé », vibrante linguale dite liquide
ز	ز	ز	ز	zâ'y	z	Z (gaz), dentale spirante sonore
س	س	س	س	sîn	s	S (sagesse), dentale spirante sourde
ش	ش	ش	ش	šîn	š	Ch (chant), palatale spirante sourde dite chuintante
ص	ص	ص	ص	ṣâd	ṣ	s emphatique, dentale spirante sourde vélarisée
ض	ض	ض	ض	ḍâd	ḍ	même valeur actuellement que ط (cf ci-dessous)
ط	ط	ط	ط	ṭâ'	ṭ	t emphatique, dentale occlusive sourde vélarisée
ظ	ظ	ظ	ظ	ẓâ'	ẓ	ḏ emphatique, interdentale spirante sonore vélarisée

ع	ع	ع	ع	ʿayn(e)	ʿ	laryngale spirante sonore
غ	غ	غ	غ	ġayn(e)	ġ	r (de la région parisienne), vélaire spirante sonore
ف	ف	ف	ف	fâʾ	f	f, labiodentale spirante sourde
ق	ق	ق	ق	qâf	q	occlusive arrière-vélaire sourde accompagné d'une explosion glottale
ك	ك	ك	ك	kâf	k	k (kaolin), palatale occlusive sourde
ل	ل	ل	ل	lâm	l	l, linguale dite liquide
م	م	م	م	mîm	m	m, labiale nasalisée
ن	ن	ن	ن	n(o)ûn	n	n, dentale nasalisée
ه	ه	ه	ه	hâʾ	h	Spirante sonore
و	و	و	و	wâw	w/û	consonne w (tramway) ou voyelle û
ي	ي	ي	ي	yâʾ	y/î	consonne y (payer) ou voyelle î longue

3. Morphologie de la langue arabe

3.1. Le nom et le nom-adjectif

3.1.1. Le genre, généralités

Du point de vue du genre, en arabe, on classe les êtres en deux catégories :

- les êtres animés dont le sexe est nettement différencié (homme/femme, chien/chienne, etc.).
- les objets abstraits ou concrets inanimés asexués (ciel, grandeur, arbre, etc.).
- les collectifs se rangent pratiquement dans l'une ou l'autre des catégories.

3.1.1.1. Genre des êtres animés sexués

Un nom singulier représentant un être animé dont le sexe est nettement différencié est masculin ou féminin selon le sexe.

Exemples 1 :

l'enfant saigne ينزف الطفلُ

la route est bloquée الطريقُ مغلقٌ

3.1.1.2. Genre des objets asexués

Un nom singulier représentant un objet concret ou abstrait inanimé et asexué est au féminin quand il finit par ة (ة).

Exemples 2 :

la pièce s'est effondrée انهارت الغرفة

Des noms bien que dépourvus de la désinence ة (ة) sont cependant féminins par nature, ce sont les mots appartenant aux catégories suivantes :

- les noms qui expriment ou rappellent l'idée de feu, de chaleur : *soleil* (شَمْسٌ), *feu* (نَارٌ), *guerre* (حَرْبٌ), *vin* (خَمْرٌ), *enfer* (جَهَنَّمَ - جَحِيمٌ) ;
- les noms du ciel et des vents : *ciel* (سَمَاءٌ), *vent* (رِيحٌ), *souffle vital* (نَفْسٌ) ;
- les noms de la terre et de la route : *route* (طَرِيقٌ), *terre* (أَرْضٌ), *l'Algérie* (الجزائر) ;
- les noms de l'habitation et de certains objets immobiliers propres au bédouin : *maison* (دَارٌ), *seau* (دَلْوٌ), *coupe* (كَأْسٌ), *bâton* (عَصَاٌ), *arc* (قَوْسٌ), *pioche* (فَأْسٌ), *marmite* (قَدْرٌ) ;
- les noms de la source : *source* (عَيْنٌ), *puits* (بَيْرٌ) ;
- les noms de certaines parties et de toutes les parties doubles du corps : *œil* (عَيْنٌ), *main* (يَدٌ), *jambe* (رِجْلٌ), *langue* (لِسَانٌ), exceptions (sont au masculin) : *joue* (خَدٌ), *sourcil* (حَاجِبٌ), *coude* (مِرْفَقٌ), *tempe* (صَدَغٌ) ;
- les noms d'animaux suivants : *hyène* (ضَبُعٌ), *lièvre* (أَرْنَبٌ), *vipère* (أَفْعَى).

Les noms suivants exigent parfois un accord masculin même si le féminin prévaut : *situation* (حَالٌ), *paix* (سَلَامٌ), *cuirasse* (دِرْعٌ), *cou* (عُنُقٌ), *foie* (كَبِدٌ), *sel* (مِلْحٌ), *couteau* (سَكِّينٌ), *scorpion* (عَقْرَبٌ), *aigle* (عَقَابٌ), *nuque* (قَفَاٌ).

Les noms singuliers représentant un objet inanimé asexué qui n'ont pas la désinence ة (ة) ou qui n'entrent pas dans la liste ci-dessus, sont masculins.

Exemples 3 :

tête (رأس), mouchoir (منديل)

3.1.1.3. Genre des collectifs

Ce genre est incertain on dit parfois qu'ils sont du genre commun, appellation qui reste équivoque. Les collectifs désignant des espèces telles que les camélidés, les ovins, les espèces chevalines mais n'ayant pas de nom d'unité, sont en général féminins.

Exemple 4 :

Les camélidés (mâles et femelles) sont rapides. الإبل سريعة

Les collectifs de plantes et d'animaux ayant un nom d'unité sont plutôt masculins :

Exemple 5 :

arbres (شجر), abeilles (نحل), etc.

3.1.1.4. Genre dans les noms-adjectifs

L'arabe tire du verbe ou du nom un certain nombre de schèmes, qui sont employés tantôt comme des noms et tantôt comme des adjectifs. Ces schèmes n'ont pas un genre qui leur appartient en propre, ils prennent celui du nom exprimé ou sous-entendu auquel ils se rapportent.

On ajoute la désinence ة (ة) et quelques fois ء ou ي pour former le féminin. Mais certains schèmes sont communs aux deux genres.

Exemples 6 :

grand كبير (fém) كبيرة

blanc أبيض (fém) بيضاء

policier شرطي (fém) شرطية

vieille/vieil (pour les deux genres) عجوز

3.1.2. Le nombre dans les noms et les noms-adjectifs

Du point de vue du nombre, l'arabe classe les noms en trois catégories :

- le singulier, un mot est au singulier quand il n'a ni la désinence du duel ni celle du pluriel externe ou quand il s'oppose à un schème dont l'usage fait son pluriel interne.

Exemples 7 :

homme رجل / hommes رجال

femme امرأة / femmes نساء

- le duel, il se forme à partir du singulier par suffixation de ان. Quand le mot se termine par la désinence ة (ة) cette lettre s'écrit ت.

Exemples 8 :

infirmier ممرض / deux infirmiers ممرضان

victime ضحية / deux victimes ضحيتان

- le pluriel externe et interne.

3.1.2.1. Le pluriel externe

Appelé aussi pluriel sain se forme à partir du singulier par suffixation de ون pour la masculin et de ات pour le féminin. Ces deux désinences se suffixent régulièrement aux participes, aux noms-adjectifs de relations, aux noms-adjectifs en فَعَّال. L'élatif muni de l'article et quelquefois les noms adjectifs en فَعِيل ont un pluriel avec ون. Les formes dérivées forment leur pluriel avec ات.

Exemples 9 :

ممرضون pl. infirmiers ممرض

ممرضات pl. infirmières ممرضة

مفقودون pl. perdus مفقود

الأكثر pl. les plus nombreux أكثر

فحوص طبية pl. examens médicaux فحص طبي

3.1.2.2. Le pluriel interne

Appelé aussi pluriel brisé, est formé sur de très nombreux schèmes (les grammairiens arabes en comptent 44) qui servent de pluriels à des singuliers dont ils diffèrent, soit par des préfixations, infixations ou suffixations, soit par des modifications de voyelles.

Exemples 10 :

نافذة fenêtré(pl) نوافذ

مدينة livre (pl) مدن

Certains noms possèdent deux pluriels, un pluriel et un pluriel du pluriel (جمع الجمع)

Exemples 11 :

شهر (mois) : أشهر (moins de 12), شهور (au delà)

Les racines quadrilitères forment leurs pluriels internes à partir des schèmes suivant :

عقارب, scorpion (pluriel : عقارب) --> فعالل

سكاكين, couteau (pluriel : سكاكين) --> فعاليل

أساتذة, professeur (pluriel : أساتذة) --> فعاللة

3.2. L'adjectif qualificatif

L'adjectif sert à caractériser le nom auquel il se rapporte. Il peut être attribut ou épithète. Seul l'accord permet de faire la différence entre les deux types d'adjectifs, le verbe *être* n'existant pas au présent (inaccompli) dans la langue arabe.

Exemple 12 :

La plaie est profonde (attribut) (الجرح عميق) / la plaie profonde (épithète) (الجرح العميق)

3.2.1. Accord de l'adjectif épithète

L'adjectif épithète s'accorde en genre, en nombre, en cas et en détermination avec un mot au singulier ou au duel ou avec un pluriel désignant des êtres humains.

Exemples 13 :

des personnes âgées أشخاص مسنّون

les deux enfants handicapés الولدان المعوّقان

L'adjectif reste le plus souvent au féminin singulier avec un sujet désignant des animaux ou des objets asexués.

Exemples 14 :

des brûlures sévères حروق خطيرة

des animaux domestiques حيوانات أليفة

L'adjectif se met généralement au féminin quand il se rapporte à un nom d'unité ou à un collectif sans nom d'unité ; on a généralement le masculin s'il se rapporte à un collectif avec un nom d'unité⁴⁹.

Exemples 15 :

un moustique dangereux بعوضة خطيرة (nom d'unité)

des gens inquiets ناس قلق/ناس قلقون (collectifs sans nom d'unité)

des moustiques porteurs du virus بعوض حامل للفيروس (collectif avec unité)

Contrairement au français, il n'est pas possible en arabe de placer l'adjectif avant le nom qu'il qualifie.

Exemple 16 :

المحاولة الأخيرة/la dernière tentative

⁴⁹ Il est important de signaler que l'accord de l'adjectif épithète avec les collectifs reste flottant comme le montre le nom *gens* qui accepte le féminin et le masculin dans l'exemple 15.

3.2.2. Accord de l'adjectif attribut

Il s'accorde comme l'adjectif épithète en genre et en nombre avec le nom auquel il se rapporte. Il peut être attribut du sujet dans une phrase nominale non régie par un verbe ou attribut d'un complément si la phrase contient un verbe d'existence ou un verbe attributif (Voir chap4/4.1 Les phrases nominales).

Exemple 17 :

une veine est obstruée وريدٌ مسدودٌ

3.3. Les nombres

On emploie deux catégories de chiffres dans le monde arabe :

- Les chiffres dit arabes (utilisés en occident et plus au Maghreb) et transmis à l'Europe par les arabes à l'époque médiévale.
- Les chiffres dit indiens utilisés au Moyen-Orient et en Egypte, ils sont postérieurs aux premiers et fonctionnent exactement de la même façon.

Arabes	0	1	2	3	4	5	6	7	8	9
indiens	٠	١	٢	٣	٤	٥	٦	٧	٨	٩

Il est important de noter que les nombres représentent un chapitre complexe de la grammaire arabe, ils sont soumis à des règles strictes et complexes, d'où notre choix d'utiliser les chiffres arabes dans la langue contrôlée (cf. utiliser 2 à la place de « deux »). (Voir chap3.III/2.2.2. Contrôle dynamique). Le choix des chiffres indiens a été écarté par souci de conformité aux normes de rédaction internationales.

3.3.1. Les cardinaux

Les cardinaux sont séparés en quatre catégories.

- Un, dans une numération, un, deux, trois, etc., l'arabe emploie أَحَدٌ, au féminin إحدى. Le chiffre وَاحِدٌ, féminin وَاحِدَةٌ est en réalité un adjectif épithète équivalent à : *seul, unique*. Sa syntaxe est celle du qualificatif. Il se décline au trois cas.

Exemples 18 :

un médicament, un seul دواءٌ واحدٌ
une seringue, une seule محقنةٌ واحدةٌ

- Deux, dans une énumération, l'arabe emploie اِثْنَانِ féminin اِثْنَانِ qui porte la flexion du duel au cas accusatif ou génitif.

Exemple 19 :

un, deux, trois etc. اِثْنَانِ أَحَدٌ،

Quand le dénombré est exprimé, on emploie le duel. اِثْنَانِ ne s'exprime que comme qualificatif corroboratif.

Exemples 20 :

deux sorties مَخْرَجَانِ
deux sorties, deux et pas plus اِثْنَانِ مَخْرَجَانِ

- De 3 à 10, le dénombré est au pluriel. Les nombres de 3 à 10 sont au masculin lorsque le dénombré est au féminin et sont au féminin lorsque le dénombré est au masculin, anomalie qui se retrouve dans d'autres langues sémitiques. Ce sont des noms. Ils ont une flexion à 3 cas :

Masculin	Féminin	
ثَلَاثَةٌ	ثَلَاثٌ	Trois
أَرْبَعَةٌ	أَرْبَعٌ	Quatre
خَمْسَةٌ	خَمْسٌ	Cinq
سِتَّةٌ	سِتٌّ	Six
سَبْعَةٌ	سَبْعٌ	Sept
ثَمَانِيَةٌ	ثَمَانٍ	Huit

تسعة	تسع	Neuf
عشرة	عشر	Dix

L'objet compté se met au pluriel au cas génitif, comme complément déterminatif du cardinal.

Exemples 21 :

trois patients ثلاثة مرضى

trois techniques ثلاث تقنيات

- De 11 à 19, l'objet compté reste au singulier. Seuls 11 et 12 s'accordent en genre avec le dénombré. De 13 à 19, la première composante du nombre est au masculin quand le dénombré est au féminin, elle est au féminin quand le dénombré est au masculin, tandis que la deuxième partie s'accorde en genre avec le dénombré. Ces nombres sont indéclinables sauf اثنا عشر et اثنتا عشر qui se déclinent en اثني عشر et اثنتي عشر au cas accusatif et génitif.

Masculin	féminin	
أحد عشر	إحدى عشرة	onze
إثنا عشر	إثنتا عشرة	Douze
ثلاثة عشر	ثلاث عشرة	Treize
أربعة عشر	أربع عشرة	Quatorze
خمسة عشر	خمس عشرة	Quinze
ستة عشر	ست عشرة	Seize
سبعة عشر	سبع عشرة	dix-sept
ثمانية عشر	ثماني عشرة	dix-huit
تسعة عشر	تسع عشرة	dix-neuf

L'objet dénombré reste au singulier et se met au cas accusatif indéterminé en fonction du spécificatif (تمييز).

Exemples 22 :

quatorze hommes أربعة عشر رجلاً

dix-neuf femmes تسع عشرة امرأة

Formation des dizaines, centaines, milliers et nombres complexes :

- Les dizaines, vingt se dit عشرون. Les autres dizaines sont les pluriels externes des unités 3 à 9. Les dizaines ont la flexion du pluriel externe masculin.

عشرون	Vingt
ثلاثون	Trente
أربعون	Quarante
خمسون	Cinquante
ستون	Soixante
سبعون	soixante-dix
ثمانون	quatre-vingts
تسعون	quatre-vingt-dix

Exemples 23 :

vingt jours عشرون يوماً

quarante minutes أربعون دقيقة

Remarque : hormis 1 et 2, les unités qui précèdent les dizaines sont au masculin quand le dénombré est au féminin, et au féminin quand il est au masculin. Les dizaines ne s'accordent pas en genre et sont jointes aux unités par و.

Exemples 24 :

quarante et une heures إحدى وأربعون ساعة

cinquante-trois minutes ثلاث وخمسون دقيقة

- Les centaines, s'écrit (مائة) ou (مئة). Pour 200, on utilise le duel. De 300 à 900, le nom du nombre qui le multiplie est sans ة et à trois flexions ; مائة reste au singulier et au cas génitif.

Les centaines sont parfois écrites en deux mots : ثلاث مائة

مئة	cent
مائتان	deux cents
ثلاثمائة	trois cents
أربعمائة	quatre cents
خمسمائة	cinq cents
ستمائة	six cents
سبعمائة	sept cents
ثمانمائة	huit cents
تسعمائة	neuf cents

L'objet compté se met au cas génitif indéterminé, comme complément déterminatif d'un état d'annexion.

Exemples 25 :

deux cents kilogrammes مائتاً كيلو غرام

trois cents mètres ثلاثمائة متر

L'ordre est : centaines – unités – dizaines.

Exemples 26 :

763 se lira alors سبعمائة و ثلاثة و ستون

- Les milliers, mille se dit ألف au pluriel آلاف. Pour 2000 on utilise le duel ألفان

Exemples 27 :

5000 خمسة آلاف

11000 أحد عشر ألفاً

30000 ثلاثون ألفاً

L'ordre est milliers - centaines – unités – dizaines

Exemples 28 :

6763 se lira ستة آلاف و سبعمائة و ثلاثة و ستون

- Les millions et les milliards,
million = مليون (pluriel : ملايين)
milliard = مليار (pluriel : ملايين)

3.3.2. Les ordinaux

Premier est bâti sur le schème أول (أولى au féminin).

Deuxième est bâti sur le schème d'un participe actif de racine défectueuse : ثانٍ (ثانية féminin).

De troisième à dixième, nous avons des schèmes فاعلة féminin فاعل, rappelant le principe actif de première forme, tirés des racines ayant formé les huit premiers nombres cardinaux. Une exception pour ستة qui donne سادس.

ثاني	deuxième
ثالث	troisième
رابع	quatrième
خامس	cinquième
سادس	sixième
سابع	septième
ثامن	huitième
تاسع	neuvième
عاشر	dixième

Tous ces ordinaux s'accordent en genre, avec une flexion comme مباني (bâtiments) pour ثاني (deuxième) et une flexion à trois cas du troisième au dixième. Ils peuvent également être définis, s'ils sont précédés de l'article الـ.

Exemple 29 :

الحادث الخامس le cinquième accident

Utilisés au cas accusatif indéfini, ils fonctionnent comme des adverbes : أولاً (premièrement).

Du onzième au dix-neuvième nous avons :

Masculin	féminin	
حادي عشر	حادية عشرة	onzième
ثاني عشر	ثانية عشرة	Douzième
ثالث عشر	ثالثة عشرة	Treizième
رابع عشر	رابعة عشرة	Quatorzième
خامس عشر	خامسة عشرة	Quinzième
سادس عشر	سادسة عشرة	Seizième
سابع عشر	سابعة عشرة	dix-septième
ثامن عشر	ثامنة عشرة	dix-huitième
تاسع عشر	تاسعة عشرة	dix-neuvième

Ces schèmes sont généralement sans flexion et s'accordent en genre avec le dénombré. S'ils sont déterminés, seul le premier élément prend l'article.

A partir du vingtième, seul le nom d'unité s'accorde en genre avec le dénombré, mais l'unité et la dizaine prennent l'article de détermination الـ.

Exemple 30 :

le vingt et unième الحادي والعشرون

Les nombres ordinaux sont des noms-adjectifs qui peuvent être employés comme nom ou comme épithète.

Exemple 31 :

le quarante-deuxième parmi eux الثاني والأربعون منهم

Au delà de cent, il faut utiliser la préposition (بعد, après).

Exemple 32 :

la 103^{ème} victime الضحية الثالثة بعد المائة

3.4. Flexion

En arabe classique, selon la fonction du nom ou de l'adjectif dans la phrase, on constate un changement dans la finale du mot. C'est ce qu'on appelle la flexion.

Celle-ci est à deux cas ou à trois cas : le nominatif, l'accusatif et le génitif.

3.4.1. Flexion à trois cas

A l'exception des mots dont la finale est *ي*, tous les singuliers et tous les pluriels internes ont une flexion de base à trois cas sans *tanwîn*, s'ils sont déterminés grammaticalement.

Ces trois flexions sont :

- le nominatif (*damma* ّ)

Exemples 33 :

l'enfant الولد

le médecin des urgences طبيب الإستعجالات

- l'accusatif (*fatha* َ)

Exemples 34 :

l'enfant الولد

le médecin des urgences طبيب الإستعجالات

- le génitif (*kasra* ِ)

Exemples 35 :

l'enfant الولد

le médecin des urgences طبيب الإستعجالات

Les noms père أب, frère أخ et possesseur ذو, ont une voyelle longue en état d'annexion.

Exemple 36 : le père de la victime

Nominatif أبو الضحية

Accusatif أباً الضحية

Génitif أبي الضحية

A l'exception des diptotes et des mots à finale ـ , tous les singuliers et tous les pluriels internes ont une flexion à trois cas avec *tanwîn* s'ils sont indéterminés grammaticalement.

Exemple 37 : enfant

Nominatif ولدٌ

Accusatif ولدًا

Cas génitif ولدٍ

Exemple 38 : habitants

Nominatif سكانٌ

Accusatif سكانًا

Génitif سكانٍ

3.4.2. Flexion à deux cas

Les flexions à deux cas affectent plusieurs catégories de mots :

3.4.2.1. Diptotes

Ces mots ont la même marque (*fatha*) au cas accusatif et génitif qui lorsque ils sont indéterminés ne prennent pas de *tanwîn*.

Parmi les diptotes on trouve :

- Des noms propres soit d'origine étrangère, soit d'origine arabe terminés par ة ou انُ auxquels s'ajoutent quelques noms d'hommes.

Exemples 39** :

L'Egypte مصرُ, Fatima فاطمةُ

En état d'annexion (complément du nom), le *noun* (ن) du duel disparaît.

Exemple 46 : les deux mains du patient

Nominatif يَدَا الْمَرِيضِ

Cas accusatif et génitif يَدَيِ الْمَرِيضِ

3.4.2.3. Flexion du pluriel externe masculin

Indéterminé ou avec l'article, le mot se termine pas (ون) au nominatif et par (ين) à l'accusatif et au génitif.

Exemple 47 : infirmiers

Nominatif مَمْرُضُونَ

Cas accusatif et génitif مَمْرُضِينَ

En état d'annexion, le *noun* (ن) du pluriel disparaît.

Exemple 48 : les infirmiers de l'hôpital

Nominatif مَمْرُضُو الْمَسْتَشْفَى

Cas accusatif et génitif مَمْرُضِي الْمَسْتَشْفَى

3.4.2.4. Flexion du pluriel externe féminin

Indéterminé (marqué par le *tanwin*), le mot se termine pas (اتُ) au nominatif et par (اتِ) à l'accusatif et au génitif.

Exemple 49 : infirmières

Nominatif مَمْرُضَاتُ

Cas accusatif et génitif مَمْرُضَاتِ

Déterminé par un article ou par annexion, le mot se termine pas (اتُ) au nominatif et par (اتِ) à l'accusatif et au génitif.

Exemple 50 : les infirmières

Nominatif الممرضات

Cas accusatif et génitif الممرضاتِ

Exemple 51 : les infirmières de l'hôpital

Nominatif ممرضاتُ المستشفى

Cas accusatif et génitif ممرضاتِ المستشفى

3.4.2.5. Flexion des mots à finale *ـ* ou *ي*

Ces mots sont des participes actifs tirés de verbes défectueux. Il faut y ajouter quelques substantifs (comme *وادي*). Concernant les flexions des pluriels quadrisyllabiques tirés de racines défectueuses, dans le cas :

Indéterminé, le (*ي*) est remplacé par le *tanwîn* au cas nominatif et génitif.

Exemple 52 : bâtiments مباني

Nominatif مباني

Cas accusatif مبانياً

Cas génitif مباني

Déterminé par article ou annexion, la marque du cas n'apparaît qu'au cas accusatif.

Exemple 53 : les bâtiments

Cas Nominatif المباني

Cas accusatif المباني

Cas génitif المباني

Exemple 54: les bâtiments de la ville

Nominatif مباني المدينة

Cas accusatif مباني المدينة

Cas génitif مباني المدينة

3.4.3. Mots sans flexion

Tous les participes passifs et les noms de lieu tirés des verbes défectueux ainsi que tous les noms terminés par *ـ* ou par *ي*, sont sans flexion ; toutefois dans la détermination ils perdent le *tanwîn*.

Exemple 55 : un bâton

Nominatif, cas accusatif, cas génitif عصاً

Exemple 56 : le bâton

Nominatif, cas accusatif, cas génitif العصا

3.4.4. Emploi des cas

Après cette présentation des différents cas, nous allons voir comment employer chacun d'entre eux.

3.4.4.1. Le cas nominatif

C'est le cas du sujet et aussi celui de l'attribut non régi par un verbe.

Exemples 57 :

les secouristes sont arrivés وَصَلَ رِجَالُ الْإِسْعَافِ
la solution (est) prête الْمَحْلُولُ جَاهِزٌ

3.4.4.2. Le cas accusatif

C'est le cas :

- du complément direct.

Exemple 58 :

informez la police أبلغوا الشرطة

- de l'attribut régi par les verbes d'existence notamment : *passer la nuit* بَيَّاتُ بَاتَ , *passer le jour* يَطْلُ ظَلَّ , *être le soir* أَمْسَى , *être dans la matinée* أَضْحَى , *être le matin* أَصْبَحَ , *rester* بَقِيَ , *demeurer* يَبْقَى , *être* كَانَ , *devenir* بَصِيرَ , *être toujours, ne pas cesser* مَازَالَ .

Exemple 59 :

si l'enfant est seul إِذَا كَانَ الطِّفْلُ وَحِيداً

- du complément circonstanciel direct.

Exemple 60 :

ne pas sortir la nuit لَا يَجِبُ الْخُرُوجُ لَيْلاً

- des mots précédés d'une particule dite du « cas accusatif » qui sont : إِنَّ (au début d'une phrase pour lui donner plus de vigueur, mais ne se traduit généralement pas), que أَنَّ, parce que لِأَنَّ, mais لَكِنَّ, peut être لَعَلَّ, ah si ... !, pourvu que لَيْتَ.

Exemple 61 :

Parce que les secouristes en auront besoin لأنَّ رجالَ الإسعاف سيحتاجون إليه

3.4.4.3. Le cas génitif

C'est le cas :

- du complément déterminatif dans l'état d'annexion.

Exemple 62 :

le casque de la victime خوذة الضحية

- du mot régi par une préposition.

Exemple 63 :

sortir de la pièce يجب الخروج من الغرفة

3.5. Le verbe⁵⁰

La conjugaison en arabe est pauvre si nous la comparons à celles des langues indo-européennes. Le verbe, en arabe classique, se présente sous deux aspects : l'accompli الماضي qui n'a qu'un mode et l'inaccompli⁵¹ المضارع qui en a trois qui diffèrent par leurs désinences. L'impératif est rattaché par les grammairiens arabe à l'inaccompli de par, à la fois, sa morphologie et la syntaxe.

⁵⁰ Nous donnons ici la conjugaison des verbes nus et nous penons la racine فَعَلَ pour illustrer les différents cas.

⁵¹ Il existe un autre mode associé à l'inaccompli qui consiste à asséner le sens du verbe, il s'agit de l'inaccompli énergique. Il est formé par l'ajout à l'inaccompli du suffixe ـَنَّ appelé *noun de corroboration* نون التوكيد. Ce mode ne se rencontre donc guère que dans les serments, les menaces et autres discours de ce genre. En français, on le rend par un futur.

Aux verbes se rattachent trois substantifs dérivés directement du verbe : un substantif abstrait ou infinitif nommé nom verbal (masdar) (Voir chap7/9.1.1.1. Le masdar équivalent de l'infinitif à valeur impérative), le participe actif (nom d'agent) et le participe passif (nom de patient).

3.5.1. Accompli

L'accompli s'obtient par la suffixation au radical du verbe, d'éléments pronominaux indiquant la personne, le genre et le nombre.

		Singulier	Duel	pluriel
1° personne	Masculin et féminin	فَعَلْتُ	Inexistant remplacée par la 1° personne du pluriel	فَعَلْنَا
2° personne	Masculin	فَعَلْتَ	فَعَلْتُمَا	فَعَلْتُمُ
	Féminin	فَعَلْتِ		فَعَلْتُنَّ
3° personne	Masculin	فَعَلَ	فَعَلَا	فَعَلُوا
	Féminin	فَعَلَتْ	فَعَلَتَا	فَعَلْنَ

Quand la dernière lettre du verbe est ن et précède l'élément pronominal de la première ou de la troisième personne du féminin pluriel, on écrit un seul ن mais avec sādā.

Exemples 64 :

nous avons contrôlé les conduites de gaz عَايَنَّا قَنَوَاتِ الْغَازِ

elles ont contrôlé les conduites de gaz عَايَنَ قَنَوَاتِ الْغَازِ

Quand cette dernière lettre est ت et précède les éléments pronominaux commençant par ت, on écrit un seul ت mais avec sādā.

Exemples 65** :

J'ai attiré l'attention لَفَتُّ الْأَنْظَارَ

L'accompli, à la première forme, se présente sous trois types :

- un type **فَعَلَ** qui est très courant et représente des verbes actifs, c'est à dire dont l'action est accomplie effectivement par le sujet.

Exemple 66 :

tousser سَعَلَ

- un type **فَعِلَ** qui se trouve assez souvent et représente des verbes actifs ou indiquant un état passager.

Exemple 67 :

tomber malade مَرَضَ

- un type **فُعِلَ** qui se rencontre assez rarement et représente des verbes exprimant un état durable (verbes qualitatifs).

Exemple 68 :

s'affaiblir ضَعُفَ

L'accompli exprime une action réalisée dans un passé vague ou considérée comme réalisée. En conséquence, selon le contexte, en français, on le rend :

- par un passé quelconque.

Exemples 69 :

La température a baissé انخفضت الحرارة

- par un présent ou un futur ou d'autres temps convenables, dans des phrases contenant une hypothèse avec **لو، إن، si**, etc.

Exemple 70 :

Vous rentrerez si les autorités vous l'autorisent ستعودون إن سمحت لكم بذلك السلطاتُ

3.5.2. Inaccompli

L'inaccompli, à la première forme, se présente sous trois types qui doivent être mis en rapport avec l'accompli. On distingue donc :

- à l'accompli en فَعَلَ correspond toujours et uniquement un inaccompli en يَفْعَلُ.

Exemple 71 :

s'affaiblir يَضْعَفُ ضَعُفَ

- à l'accompli en فَعِلَ correspond presque toujours un inaccompli en يَفْعِلُ.

Exemple 72 :

tomber malade يَمْرُضُ مَرَضَ

- à l'accompli en فَعَلَّ correspond un inaccompli en يَفْعِلُ، يَفْعُلُ، يَفْعُلُّ. Seul l'usage permet de préciser le timbre de cette voyelle.

Exemples 73 :

couper قَطَعَ، يَقْطَعُ

descendre نَزَلَ، يَنْزِلُ

tomber سَقَطَ ، يَسْقُطُ

L'inaccompli a plusieurs modes : l'indicatif, le subjonctif, l'apocopé. Ces modes diffèrent, en général, par la terminaison.

3.5.2.1. Inaccompli indicatif

L'inaccompli indicatif s'obtient essentiellement par la préfixation d'éléments pronominaux indiquant la personne, et secondairement, par l'emploi de suffixes indiquant le mode, le genre et le nombre.

		Singulier	Duel	pluriel
1° personne	Masculin et féminin	أَفْعَلُ	Inexistant remplacée par la 1° personne du pluriel	تَفْعُلُ
2° personne	Masculin	تَفْعَلُ	تَفْعَلَانِ	تَفْعَلُونَ
	Féminin	تَفْعَلِينَ		تَفْعَلْنَ
3° personne	Masculin	يَفْعَلُ	يَفْعَلَانِ	يَفْعَلُونَ
	Féminin	تَفْعَلُ	تَفْعَلَانِ	يَفْعَلْنَ

Quand la dernière lettre du verbe est un ن, on écrit نٌ, aux deuxième et troisième personnes du féminin pluriel : تُعَايِنُنَّ (elles contrôlent) تُعَايِنَنَّ (vous contrôlez) pour يُعَايِنَنَّ

L'inaccompli indicatif exprime une action qui dure ou qui se répète ou qui est en voie de réalisation. En conséquence, en français, on le rend selon le contexte, par :

- l'imparfait, surtout quand il est précédé du verbe كان à l'accompli.

Exemple 74 :

il ne respirait pas كان لا يَتَنَفَّسُ

- le présent.

Exemple 75 :

la fumée envahit la pièce الدخانُ يجتاح الغرفة

- le futur, surtout s'il est précédé de سوف ou سَ.

Exemple 76 :

quand vous entendrez le signal d'alarme عندما تسمعون جهاز الإنذار

3.5.2.2. Inaccompli subjonctif

L'inaccompli subjonctif diffère de l'indicatif en ce qu'il a un son َ (fatha) au lieu de ُ (damma) à certaines personnes, et en ce qu'il perd le ن final dans les désinences contenant une voyelle longue.

		Singulier	duel	pluriel
1° personne	Masculin et féminin	أَفْعَلْ	Inexistant remplacée par la 1° personne du pluriel	نَفْعَلْ
2° personne	Masculin	تَفْعَلْ	تَفْعَلَا	تَفْعَلُوا
	Féminin	تَفْعَلِي		تَفْعَلْنَ
3° personne	Masculin	يَفْعَلْ	يَفْعَلَا	يَفْعَلُوا
	Féminin	تَفْعَلْ	تَفْعَلَا	يَفْعَلْنَ

D'une façon générale, le subjonctif exprime une action à réaliser, subordonnée à une autre. Il ne correspond pas toujours au subjonctif français. On l'emploie dans des propositions introduites :

- par *أَنَّ* *que* ou des locutions où l'on trouve cette particule telles que : *avant que* *يُقبل أن يتوقف*, *après que* *يُبعد أن*.

Exemple 77 :

avant que le saignement s'arrête *يُقبل أن يتوقف النزيف*

- par *لِكَيْ*, *لِكَيْ*, *afin que*, *حتى* *jusqu'à ce que*, *فَ* *en sorte que*.

Exemple 78 :

afin que le blessé reste calme *حتى يبقى الجريح هادئاً*

- Après *لَنْ* *ne ... pas*, on a toutefois une proposition principale avec un sens du futur.

Exemple 79 :

Il ne mettra pas longtemps à éteindre l'incendie *لَنْ يستغرق وقتاً طويلاً لإخماد الحريق*

Quand la proposition introduite par *حتى* ou *أَنَّ* ou les locutions formées de cette particule, indique une action déjà réalisée, l'arabe laisse le verbe à l'accompli.

Exemple 80 :

Il mit le pansement sur la plaie après l'avoir nettoyée *ضمّد الجرحَ بعد أن نَقَّاهُ*

3.5.2.3. Inaccompli apocopé

Identique à celle du subjonctif sauf aux personnes où celui-ci a un son *ˤ* (*fatha*). A ces personnes, l'apocopé a une finale *ˤ* (*sukûn*) et au lieu de se terminer par deux syllabes ouvertes, il se termine sur une syllabe fermée, d'où son nom d'apocopé.

		Singulier	Duel	pluriel
1° personne	Masculin et féminin	أَفْعَلْ	Inexistant remplacée par la 1° personne du pluriel	نَفْعَلْ
2° personne	Masculin	تَفْعَلْ	تَفْعَلَا	تَفْعَلُوا
	Féminin	تَفْعَلِي		تَفْعَلْنَ
3° personne	Masculin	يَفْعَلْ	يَفْعَلَا	يَفْعَلُوا
	Féminin	تَفْعَلْ	تَفْعَلَا	يَفْعَلْنَ

Un verbe se lit à l'apocopé quand il est précédé des particules :

- *que* فَاَ qui se joint au verbe et exprime un ordre (ne pas confondre avec la particule du subjonctif لِـ *afin que*).

Exemple 81** :

ليرحل، فليرحلْ qu'il parte

- *ne ... pas* لَمْ qui donne à l'inaccompli un sens passé.

Exemple 82 :

لم يتوقف النزيف le saignement ne s'est pas arrêté

- *ne ... pas encore* لَمْ، qui donne aussi un sens passé à l'inaccompli (ne doit pas être confondue avec la particule *quand* لَمَّا، qui se trouve seulement dans l'accompli).

Exemple 83 :

لَمْ يَصِلُوا ils ne sont pas encore arrivés

3.5.3. Impératif

L'impératif n'existe qu'à la deuxième personne, celui-ci diffère seulement de l'inaccompli apocopé par la chute de l'élément pronominal ت. A la suite de cette chute, il reste un groupe imprononçable en arabe classique.

On préfixe donc une voyelle avec attaque vocalique **ا** ou bien **إ** (*alif* prosthétique) qui forme à nouveau une syllabe fermée. On a **ا** avec les types **يَفْعُلْ** et **يَفْعِلْ**. On a **إ** avec **يَفْعُلْ**. Cette attaque vocalique est instable.

Quand l'impératif n'est pas exactement au début du discours, elle tombe et on écrit :
Exemples 84 :

appelle les secours **فَاتَّصِلْ بِالْإِسْعَافِ**
puis, coupez le câble **ثُمَّ اقْطَعُوا السَّلْكَ**

L'injonctif aux première et troisième personnes se rend en arabe par la personne correspondante de l'inaccompli apocopé précédé de **لِ** *que* ou bien de **فَ** *il faut que*.

Exemples 85** :

sortons ! **فَانْخَرِجْ**
qu'ils sortent ! **لِيَخْرُجُوا**

En arabe, la négation ne peut pas être utilisée avec l'impératif, par conséquent une interdiction se rend par l'inaccompli apocopé précédé de **لَا** *ne... pas*.

Exemple 86 :

ne déshabillez pas la victime **لَا تَخْلَعُوا مَلَابِسَ الضَّحِيَّةِ**

3.6. Les prépositions

3.6.1. Généralités

Les prépositions gouvernent toutes le cas génitif. Certaines d'entre elles sont clitiques et s'accrochent donc au mot qui suit. Elles expriment des notions : de mouvement en avant, de recul, de position, d'accompagnement, etc. Ces notions peuvent prendre des nuances assez diverses, de sorte qu'une même préposition française peut se rendre par différentes prépositions arabes. Le rôle des prépositions est souvent fort grand. Un même verbe peut exprimer en effet des sens différents, selon la préposition qui introduit l'objet. (Voir chap7/4. Les prépositions liées aux verbes dans le modèle syntaxico-sémantique TACCT)

3.6.2. La préposition ب

Elle se lie au mot qu'elle introduit et :

- exprime une idée de contiguïté temporelle ou spatiale (presque synonyme de في dans ce cas).

Exemples 87 :

بالمدينة dans la ville

بالنَّهار pendant la journée

- modifie le sens de certains verbes :

Exemples 88 :

قام (se lever), قام بـ (entreprendre, réaliser)

سمع (entendre), سمع بـ (entendre parler de)

- le moyen ou l'instrument

Exemple 89 :

أحمد النَّارَ بمنشفة مبلَّلة il a éteint le feu avec un linge mouillé

3.6.3. La préposition في

Elle devient فيَّ avec le pronom affixe de la première personne du singulier et marque :

- le lieu où l'on est (*dans, à*)

Exemple 90 :

الضحية في سيارتها la victime est dans sa voiture

- le temps, la situation où l'on se trouve.

Exemple 91 :

إذا كنتم في وضعية حرجة si vous êtes dans une situation critique

- l'objet que l'on considère, qu'on étudie.

Exemple 92 :

مختص في الإسعاف spécialisé dans le secourisme

3.6.4. La préposition منْ

Elle devient مَنْ devant un أَ et مَنَّاْ avec les pronoms affixes de la première personne et elle indique :

- le point de départ dans le temps et l'espace.

Exemples 93 :

إذا عدتم من بلد مصاب بالوباء si vous revenez d'un pays touché par l'épidémie
من البداية إلى النهاية depuis le début jusqu'à la fin

- le lieu par où l'on passe.

Exemple 94 :

أخرج من النافذة sors par la fenêtre

- l'origine, la cause ou la matière

Exemples 95 :

من أوروبا d'Europe
قفازات من المطاط des gants en caoutchouc

- introduit le deuxième terme d'une comparaison

Exemple 96 :

هذه الطريقة أسهل من الطريقة السابقة Cette méthode est plus facile que la précédente

3.6.5. La préposition عَنْ

Elle devient عَنْْ devant أَ et عَنَّْاْ avec les pronoms affixes de la première personne. Elle marque :

- la séparation, l'éloignement.

Exemple 97 :

يجب الابتعاد عن النار s'éloigner du feu

- la défense et la préservation

Exemple 98** :

يدافع عن حقوقه il défend ses droits

- le rapport

Exemple 99 :

عن يمينكم à votre droite

3.6.6. La préposition إلى

Elle devient إِلَيّْ avec les pronoms affixes. Elle marque le mouvement vers, au sens propre et figuré :

Exemple 100 :

يجب الذهاب إلى المستشفى se rendre à l'hôpital

3.6.7. La préposition لِ

Elle devient لِ devant les pronoms affixes sauf celui de la première personne et elle se lie au mot qui suit. Elle marque :

- la cause ou le but (*pour, à cause de, du fait de, etc.*)

Exemple 101 :

لإخماد النيران pour éteindre le feu

- la possession

Exemple 102 :

له كسرٌ مفتوح il a une fracture ouverte

- la dépendance ou la relation (quand l'annexion est impossible).

Exemple 103 :

المعهد العالي للبحوث والدراسات الزلزالية
l'institut supérieur des recherches et des études sismologiques

3.6.8. La préposition على

Elle devient عَلَىّ avec un pronom affixe. Elle exprime :

- un emplacement

Exemple 104 :

على الرّصيف sur le trottoir

- une obligation

Exemple 105 :

عليكم بإخلاء المكان Vous devez quitter les lieux

3.6.9. La préposition كـ

Elle se lie au mot qui suit et marque la similitude. Elle ne peut être suivie d'un pronom.

Exemple 106 :

Comme suivant كالتالي

3.6.10. La préposition حتّى

Elle ne peut pas être suivie d'un pronom affixe mais peut être suivie d'un pronom isolé (حتى أنا) elle exprime :

- Une limite dans le temps.

Exemple 107 :

jusqu'à la fin du traitement حتى نهاية العلاج

- Une inclusion (*y compris, même*)

Exemple 108 :

y compris les fenêtres حتى النوافذ

3.6.11. La préposition مع

Elle indique :

- l'accompagnement dans le temps et dans l'espace

Exemples 109 :

à l'aube مع الفجر

en contact avec une personne atteinte du virus على اتصال مع شخص مصاب بالفيروس

- l'opposition

Exemple 110 :

malgré cela مع ذلك

- la possession

Exemple 111 :

vous avez sur vous une trousse de soin معكم حقيبة إسعاف

3.6.12. La préposition **عند**

Elle indique:

- la contiguïté dans l'espace et le temps. (*auprès de, chez, à côté de, etc.*)

Exemples 112 :

chez le médecin عند الطبيب

au matin عند الصباح

- possession

Exemple 113 :

il a de l'expérience عنده خبرة

3.7. Les pronoms

L'arabe a deux séries de pronoms personnels : les pronoms isolés et les affixes.

3.7 1. Pronoms personnels isolés

Ils sont au nombre de douze :

		singulier	duel	pluriel
1° personne	fém et masc	أنا	manque	نحن
2° personne	masculin	أنتَ	أنثما	أنتم
	féminin	أنتِ		أنتن
3° personne	masculin	هو	هما	هم
	féminin	هي		هن

Il n'y a pas de distinction entre le féminin et le masculin à la première personne ainsi qu'au duel. Pour le duel et le pluriel de la première personne on utilise le même pronom (نحن). Devant un *alif* (ا) ayant perdu son attaque vocalique, les pronoms أنتم et هم deviennent أنتم et هم avec une voyelle de liaison (*damma* ُ).

Exemple 114 :

ce sont eux les professionnels هم المحترفون

Ces pronoms isolés s'emploient :

- pour attirer l'attention sur le sujet du verbe et quelquefois sur un autre pronom complément.

Exemple 115 :

c'est elle qui donne l'alerte هي من يعطي الإنذار

- comme sujet d'une phrase nominale.

Exemple 116 :

c'est elle la plus efficace هي الأكثر فعالية

- comme attribut dans une phrase nominale interrogative.

Exemple 117 :

quelle est-elle ? ما هي ؟

- comme disjonctif entre le sujet d'une phrase nominale et l'attribut déterminé.

Exemple 118 :

C'est le chauffeur, le responsable إنَّ السائق هو المسؤول

3.7 2. Pronoms personnels affixes

Ils sont au nombre de douze :

		singulier	duel	Pluriel
1° personne	féminin et masculin	ي	manque	نا
2° personne	masculin	كَ	كُما	كُم
	féminin	كِ		كُنَّ
3° personne	masculin	هُ	هُما	هُم
	féminin	هَآ		هُنَّ

Ces pronom peuvent être suffixés soit à un nom (صدره son thorax) pour exprimer la possession, soit à un verbe (نَوَّمَهُ il l'a endormi), soit à une préposition (مَعَهُ avec lui).

Les pronoms affixes peuvent occuper différentes fonction dans une phrase :

- Complément d'objet direct

Exemple119 :

أَخَذَهُ il l'a pris

- Complément d'objet indirect

Exemple120 :

إِذَا رَأَيْتُمُوهُ si vous l'avez vu.

- Complément du nom

Exemple121 :

إِذَا دَمَّرَ الزَّلْزَالُ مَنْزِلَكُمْ si le séisme a détruit votre maison

Après un verbe à la deuxième personne du masculin pluriel de l'accompli, on ajoute un و pour la liaison.

Exemple 122 :

جَرَّبْتُمُوهَا = جَرَّبْتُمْ + هَا vous l'(elle) avez essayée

Après un verbe se terminant par وا (la troisième personne du masculin pluriel de l'accompli) ou aux deuxième et troisième personnes du masculin pluriel du subjonctif ou de l'apocopé, l'alif (ا) orthographique tombe.

Exemples 123 :

طَلَبُوهُ ils l'avaient demandé

حَتَّى يَسْعَفُوكُمْ jusqu'à ce qu'ils vous secourent

لَمْ يَتَوَقَّعُوهَا ils ne l'(elle) ont pas prévue

Les pronoms affixes de la troisième personne, sauf هَا, prennent une *kasra* (ـِ) s'ils sont précédés eux-mêmes d'une *kasra* ou de ي portant *sukûn* (ـِ).

Exemples 124 :

بِهِ avec lui

سَقَطَتْ عَلَيْهِمْ elle est tombée sur eux

Le pronom affixe est **ني** à la première personne du singulier quand il complète un verbe. Avec ce pronom, les prépositions **من** et **عن** redoublent le **ن**.

Exemples 125** :

il m'a trouvé **وجدني**

منّي = من + ني (de moi)

Après les prépositions **في**, **على**, **لدى**, le pronom **ي** devenu **يَ**, se confond dans l'écriture avec la voyelle de la préposition. Il en va de même avec les noms au cas accusatif et génitif du duel.

Exemples 126 :

sur moi **عليّ**

mes deux mains **يديّ**

Devant un alif (ا) ayant perdu l'attaque vocalique, les pronoms **كُم** et **هُم** deviennent **كُ** et **هُ**.

Exemple 127 :

la tempête vous a frappés **ضربتكم العاصفة**

4. Syntaxe de la langue arabe

4.1. Les phrases nominales

4.1.1. Valeur et structure

La phrase nominale exprime une constatation, une définition, non située dans le temps. Elle se compose schématiquement :

- d'un sujet (مبتدأ) au nominatif ou au cas accusatif s'il est précédé d'une particule dite « du cas accusatif ».
- d'un attribut au nominatif (خبر), cet attribut peut être déterminé ou indéterminé.

La phrase nominale se rend en français à l'aide du verbe *être* au présent.

Exemple 128 :

le danger est imminent الخطر وشيكٌ

La phrase nominale minimale peut se réduire à deux termes qui peuvent être :

- un pronom + un nom, exemple 137 : هو طبيب مختص (il est médecin spécialiste)
- un nom + un nom, exemple 138 : الدخان خطرٌ (la fumée est un danger)
- un nom + un adjectif, exemple 139 : الدخان كثيفٌ (la fumée est épaisse)
- un démonstratif + un nom, exemple 140 : هذه الحالة (ce cas)
- un nom + un verbe, exemple 141 : الجهاز يتوقف بشكل أوتوماتيكي (l'appareil s'arrête automatiquement)

4.1.2. Accord de l'attribut

L'attribut s'accorde en genre et en nombre avec les mots au singulier et au duel ou avec un pluriel désignant des êtres humains.

Exemples 129 :

l'enfant [est] inconscient الطفل فاقدٌ للوعي

les deux personnes [sont] malades الشخصان مريضان

les blessés [sont] nombreux الجرحى كثيرون

L'attribut reste le plus souvent au féminin singulier avec un sujet désignant des animaux ou des objets asexués.

Exemples 130 :

la température [est] élevée الحرارة مرتفعة

les scorpions [sont] venimeux العقارب سامة

L'attribut se met généralement au féminin quand il se rapporte à un nom d'unité ou à un collectif sans nom d'unité, on a généralement le masculin s'il se rapporte à un collectif avec un nom d'unité.

Exemple 131 :

le moustique [est] vecteur de maladies graves البعوضة حاملةً لأمراض خطيرة

Avec deux sujets, l'attribut se met au duel.

Exemple 132 :

les deux veines [sont] sont obstruées الوريدان مسدودان

Avec trois sujets représentant des êtres humains, on a un accord du pluriel. Le masculin l'emporte sur le féminin.

Exemple 133 :

le médecin, l'infirmière et la famille [sont] présents الطبيب و الممرضة و العائلة حاضرون

4.1.3. Les complexes à l'intérieur du syntagme

4.1.3.1. L'apposition

L'apposition est une information facultative se rapportant à un élément de la phrase. Le groupe oppositionnel se place en arabe généralement avant l'élément auquel il se rapporte. Il n'y pas de virgule entre les deux. La juxtaposition peut être entre (nom + nom, nom + épithète, nom + adjectif relatif, démonstratif + nom) :

- nom + nom

Exemple 134** :

L'homme, ton frère, est parti رجل الرجل أخوك

- nom + adjectif

Exemple 135 :

seringue de petite taille محقنة صغيرة الحجم

- nom + adjectif relatif

Exemple 136 :

le défibrillateur que tu utilises est automatique. مزيل الرجفان الذي تستعمله آلي.

- nom + démonstratif

Exemple 137** :

celui-ci, cet enfant هذا الطفل

4.1.3.2. L'utilisation des démonstratifs

Les démonstratifs en arabe paraissent sous deux aspects. Les uns sont des démonstratifs de proximité. Ce sont :

		masculin	féminin
Singulier		هذا	هذه
Duel	nominatif	هذان	هتان
	cas accusatif et indirect	هذين	هتين
Pluriel		هؤلاء	

Les autres sont des démonstratifs d'éloignement. Ce sont :

		masculin	féminin
Singulier		ذلك	تلك
Duel	nominatif	ذانك	تانك
	cas accusatif et indirect	ذينك	تينك
Pluriel		أولئك	

L'arabe emploie ces démonstratifs toujours comme des pronoms :

- soit seul,

Exemple 138 :

utilisez celle-ci استعملوا هذه

- soit comme appositif d'un nom propre ou d'un état d'annexion,

Exemple 139** :

est-ce que la victime est cette personne ? هل الضحية الشخص هذا ؟

- soit suivi d'un nom ou d'un adjectif qui lui sert d'appositif et prend toujours l'article. Il est rendu en français par les adjectifs démonstratifs *ce, cet, ces...*

Exemples 140** :

celui-ci, cet homme (cet homme-ci) الرَّجُلُ هذا

ceux-là, ces hommes (ces hommes) أولئك الرجال

4.1.3.3. La détermination et l'indétermination

4.1.3.3.1. Indétermination

Un mot est indéterminé quand il est employé seul. Le français le traduit par un correspondant précédé de l'article indéfini *un, une, des*, sauf au duel et dans les nombres.

Exemples 141 :

un accident حادث

deux options خياران

des médecins أطباء

trois jours ثلاثة أيام

4.1.3.3.2. Détermination

Un mot est déterminé grammaticalement quand il est précédé de l'article ou en état d'annexion. Il n'a pas alors de *tanwîn*. Il faut distinguer la détermination grammaticale de la détermination sémantique qui appartient au sens du mot. Par exemple مُحَمَّدٌ (*Muhammad*) est indéterminé grammaticalement (il n'a pas d'article et n'est pas en état d'annexion, mais il est déterminé sémantiquement car c'est un personnage connu.

4.1.3.3.2.1. Détermination par l'article الـ

L'article الـ est un article unique invariable, il détermine n'importe quel nom commun sans distinction de genre ou de nombre.

Exemples 142 :

la police الشرطة

les personnes الأشخاص

L'article الـ peut avoir une valeur démonstrative

Exemple 143** :

ce jour-ci اليوم

4.1.3.3.2.2. Détermination par annexion

L'état d'annexion présente deux valeurs :

- ce peut être une annexion de dépendance, d'appartenance. Il comprend alors deux termes ou davantage dont le premier détermine le deuxième, le deuxième le troisième et ainsi de suite. Seul le dernier terme prend l'article si le sens l'exige.

Exemple 144 :

[les] deux jambes [de] la victime سَاقَي الضحية

[les] affaires [de] la victime أَغْرَاضُ الضحية

[les] services [de l'] hôpital [de] la ville مَصَالِحُ مَسْتَشْفَى الْمَدِينَةِ

- ce peut être une annexion de qualification composée uniquement de deux termes. Le premier est une épithète, le second un nom muni de l'article.

Exemple 145 :

une plaie petite [de] taille جَرَحٌ صَغِيرٌ الْحَجْمِ

Si l'épithète se rapporte à un nom déterminé, elle prend l'article.

Exemple 146 :

la plaie petite [de] taille الْجَرَحُ الصَّغِيرُ الْحَجْمِ

Remarque : Dans les deux cas, l'accord de l'épithète se fait comme d'ordinaire, en genre, en nombre et en cas avec le mot qualifié. (Voir chap7/11.1. Syntagmes nominaux N1-de-N2)

4.1.3.4. L'élatif

Le schème de l'élatif est أَفْعَلٌ. L'arabe l'utilise pour marquer une comparaison entre deux êtres ou deux objets. Le schème est invariable en genre et en nombre et la comparaison est introduite par مِنْ *par rapport à*, rendu en français par *que*. L'élatif est dit diptote parce qu'il ne prend que deux flexions selon le cas et jamais de *tanwîn*. (Voir chap4/3.4.2. Flexion à deux cas)

Exemples 147 :

cette substance liquide est plus dangereuse que l'autre هذه المادة السائلة أخطر من الأخرى
ces deux tests sont plus précis que les autres هذان الإختباران أدق من الآخرين

4.1.3.5. Le superlatif

Le schème أَفْعَلٌ sert aussi de superlatif. On l'emploie :

- muni de l'article, avec accord, comme pour une épithète.

Exemple 148 :

les plus grands risques المخاطر الكبرى

- comme premier terme d'un état d'annexion mais invariable en genre et en nombre.

Exemple149 :

elle est parmi les plus difficiles opérations ٥ هي من أصعب العمليات

4.2. Les phrases verbales

4.2.1. Valeur et structure

La phrase verbale exprime une action accomplie ou en voie d'accomplissement. Elle peut se réduire à un verbe, accompagné de ses éléments pronominaux et de ses désinences.

Exemples 150 :

vous avez ressenti une douleur شعرتم بألم

vous ressentez une douleur تشعرون بألم

Elle peut comprendre aussi un verbe avec un sujet lexical et l'accord du verbe est différent selon que le sujet suit ou précède.

4.2.2. Accord du verbe

L'accord du verbe se fait selon la nature et la place du sujet auquel il se rapporte :

- Si le sujet est placé après le verbe et désigne un animé, l'accord se fait uniquement en genre jamais en nombre.

Exemples 151 :

la victime respire تتنفس الضحية

les victimes/les deux victimes respirent تتنفس الضحايا/الضحيّتان

le blessé respire يتنفس الجريح

les blessés/les deux blessés respirent يتنفس الجرحى/الجريكان

Remarque : si le sujet est un inanimé au pluriel, le verbe se met au féminin singulier.

Exemple 152 :

انهارت المنازل les maisons se sont effondrées

- Si le sujet est placé avant le verbe, l'accord se fait en genre et en nombre avec le sujet, sauf si celui-ci est un pluriel inanimé, auquel cas il se met au féminin singulier.

Exemples 153 :

la victime respire الضحية تتنفسُ

les victimes/les deux victimes respirent الضحايا يتنفسون/الضحيتان تتنفسان

le blessé respire الجريح يتنفس

les blessés/les deux blessés respirent الجرحى يتنفسون/الجريحان يتنفسان

les maisons s'effondrent المنازل تنهار

Souvent cet ordre sujet + verbe est amené par la présence d'une particule dite du « cas accusatif » en tête de phrase.

Exemple 154 :

les pompiers sont arrivés إن رجال الإطفاء وصلوا

4.3. Ordre des mots

L'ordre des mots, en arabe classique, est relativement variable.

D'une manière générale, on met au début de la phrase le mot sur lequel on veut attirer l'attention et l'on termine sur le terme le plus long ou le plus riche de sens ou de sonorité.

En phrase nominale, l'ordre habituel est : sujet + attribut + complément.

Exemple 155 :

Les secouristes sont en mesure de vous aider رجال الإسعاف قادرون على مساعدتكم

Mais l'attribut gouverné par une préposition peut précéder le sujet.

Exemple 156 :

un blessé [est] dans la voiture في السيارة جريحٌ

En phrase verbale, l'ordre habituel est : verbe + sujet + complément direct + autres compléments s'ils sont longs.

Exemple 157 :

Le médecin a fait une injection à l'enfant أعطى الطبيبُ الطفلَ حقنةً

Quand on a, au début de la phrase verbale, une particule dite du « cas accusatif », le sujet vient toujours avant le verbe.

Exemple 158 :

La victime souffre d'un mal de tête إنَّ الضحيةَ تعاني من ألم في الرأس

Quand il n'y a pas de nom-sujet exprimé, on fait suivre la particule du cas accusatif d'un pronom affixe du genre, du nombre et de la personne du verbe.

Exemple 159 :

tu feras cela en cas d'accident إنَّكَ ستفعلُ ذلك في حال حادث

5. Conclusion

Nous avons voulu, en abordant sommairement les principaux microsystèmes constituant le système général de la langue arabe, montrer dans ce chapitre combien le fonctionnement de la langue arabe est différent de celui de la langue française. Nous nous sommes contentés de présenter les éléments clés de la langue qui ont trait directement ou indirectement à la traduction automatique. En partant de la morphologie à la syntaxe, nous avons démontré avec une analyse microsystemique que le système linguistique arabe présente des particularités qui lui sont inhérentes et qui font sa richesse. Cette analyse est conduite dans un environnement monolingue mais aussi contrastif à celui de la langue française, ce qui confère à ce chapitre une place centrale dans ce travail de recherche. La langue contrôlée LiSe, conçue au terme de nos travaux, se base sur les conclusions de l'analyse microsystemique des deux systèmes linguistiques français et arabe pour constituer ses règles. Les textes non contrôlés sont soumis à des contraintes provenant de plusieurs types de contrôle différents qui ne perdent pas de vue les caractéristiques de chacune des langues source et cible.

Ce chapitre a permis aussi de faire la lumière sur des problèmes divers mais qui n'ont pas exactement la même ampleur ni la même nature dans chacune des langues. En d'autres termes, certaines divergences peuvent être plus importantes que d'autres, nécessitant des règles de transfert lourdes, comme nous l'avons montré dans la partie traitant les numéraux. L'identification de ces écarts ainsi que le degré de leur complexité permet la mise en place de la langue contrôlée, qui a pour mission non seulement d'accroître la compréhensibilité des textes, mais aussi l'allègement du système de traduction automatique. Elle permet également en saisissant les subtilités du français et de l'arabe, la création d'un modèle de traduction automatique isomorphique. Le système de traduction TACCT est un système de transfert à base de règles linguistiques qui prend en compte le fonctionnement de chaque microsystème mentionné dans ce chapitre ainsi que les interrelations entre les différentes composantes. Ceci, en gardant une approche contrastive avec le français dans l'objectif de réduire les écarts et de rapprocher deux systèmes linguistiques à priori éloignés.

Chapitre 5

TRADUCTION AUTOMATIQUE

1. Introduction

La traduction automatique appartient au vaste domaine de Traitement Automatique des Langues, qui va de la recherche d'informations dans les textes à la traduction automatique, en passant par les systèmes de communication homme-machines, les vérificateurs orthographiques et grammaticaux, mais aussi les dictionnaires électroniques. Elle occupe aujourd'hui une place prépondérante dans ce vaste domaine compte tenu des enjeux stratégiques et économiques qu'elle présente. Pour arriver à son état actuel, la traduction automatique a connu une histoire tumultueuse : les grandes étapes seront retracées dans ce chapitre.

Dans ce chapitre, nous allons voir dans un premier temps un aperçu historique avec les trois grandes étapes de l'histoire de la traduction automatique présentant les débuts, les désenchantements et l'état actuel du domaine de la traduction automatique. Dans un second temps, nous nous plongerons dans les différentes architectures des systèmes de traduction.

2. Histoire de la traduction automatique

2.1. De l'engouement des débuts à la désillusion du rapport ALPAC

C'est en 1946, aux Etats-Unis, que Weaver propose d'utiliser le déchiffrement cryptographique afin de traduire des textes de façon automatique. En 1952, au MIT, la première conférence sur la traduction automatique signe les réels débuts des recherches dans le domaine. Principalement à cause du contexte économique-politique de la guerre froide, les premiers essais furent accueillis avec un enthousiasme indéniable, en témoignent les premiers systèmes comme le G.A.T. « Georgetown Automatic Translation », élaboré par l'équipe de Dorstert et Mc Donald. Ces derniers étaient en effet très largement affectés à la traduction du russe à l'anglais dans le domaine scientifique.

Malheureusement, les espoirs suscités par ces débuts fulgurants furent rapidement déçus. En effet, bon nombre de difficultés théoriques avaient alors été sous-estimées. C'est ainsi qu'au début des années 60, une certaine désillusion succéda à l'engouement des premières recherches. La nature des systèmes, qualifiés par la suite de « système de première génération », explique en partie cette désillusion. Ils effectuaient des traductions mot à mot à partir de dictionnaires sans s'occuper des représentations syntaxiques des phrases.

Contrairement aux travaux américains, les recherches en URSS avaient pris en compte quand à elles ces aspects plus théoriques. Celles-ci intégraient des questions d'analyses et se dirigeaient déjà vers l'élaboration d'un « langage intermédiaire universel ». Ce langage est constitué de modélisations logico-mathématiques. Les travaux de l'équipe de Rosenzweig à Moscou auxquels ont participé des linguistes réputés tels que Mel'cuk préfiguraient ce type de systèmes dit de « deuxième génération ».

En 1965, la commission ALPAC, Automatic Language Processing Advisory Committee, présente un rapport. Ce dernier, commandé par les agences gouvernementales aux Etats-Unis, fait suite aux premiers constats de Bar Hillel en 1960. Le rapport a pointé plusieurs difficultés. Il critiquait notamment la qualité et le coût des traductions, celles-ci nécessitant d'importantes révisions par l'humain. Ce rapport ne prouvait pas que le projet n'aboutirait pas mais il montrait plus pragmatiquement les lacunes restant à combler.

Les conséquences de ce dramatique rapport ne se font pas attendre et les crédits aux équipes de recherche sont alors considérablement réduits. Seuls restent (ou presque) les financements militaires servant évidemment des fins militaires et qui s'attachaient à étudier les modalités et problèmes de transfert du russe à l'anglais, en se concentrant sur les textes scientifiques et techniques (espionnage militaire et industriel).

2.2. De 1965 à 1975 : un domaine en berne

Le déclin brutal des financements n'a pas stoppé complètement les recherches dans le domaine de la traduction automatique mais a cependant largement réduit le nombre et l'importance des groupes de recherche investis dans ce domaine. Les dix ans suivant le rapport ALPAC, il en fut ainsi. Certaines équipes ont cependant réussi à continuer leurs travaux de recherche, on peut ainsi mentionner : à l'université du Texas, les travaux sur le système METAL et ceux de l'équipe Latsec Inc sur le système SYSTRAN ; à Montréal les travaux sur TAUM et les recherches de Vauquois à Grenoble au CETA (qui deviendra en 76 le GETA). La plupart de ces systèmes s'orientaient déjà vers les systèmes dits de « deuxième génération ».

2.3. De 1975 à nos jours : le renouveau

Dès 1975, on constate un véritable regain de l'intérêt pour les travaux dans le domaine de la traduction automatique. Cette reprise a deux causes principales. Elle s'explique à la fois par un développement des besoins des entreprises privées et publiques et par conséquent par un accroissement du marché et donc des sources de financement. Mais également par l'amélioration profonde et continue des outils informatiques, tant sur le plan technique (accroissement de la puissance des ordinateurs), que sur le plan théorique (nouveaux langages de programmation, modélisation plus poussée).

Les nouveaux systèmes de « deuxième génération » sont plus perfectionnés que les anciens systèmes de « première génération ». Certains sont mêmes capables d'un autocontrôle, grâce à des langages informatiques évolués qui permettent de faire appel de façon récursive et contrôlée à des sous-grammaires, ils peuvent détecter un cheminement erroné et revenir en arrière pour chercher un nouveau cheminement, et ainsi de suite. Parmi ces systèmes, on peut citer : ARIANE (GETA, Grenoble, russe-français, anglais-malais, français-allemand, et d'autres), METAL (Siemens, allemand-

anglais, allemand-espagnol), TITUS (Institut du Textile de France, français, anglais, allemand et espagnol).

Le regain d'intérêt pour le domaine de la traduction automatique s'est encore accentué dans les dix années qui ont suivi. En France, de 83 à 87, est lancé le projet national de traduction assistée par ordinateur PN-TAO piloté par l'Agence pour l'Informatique. Au Canada, il y a eu le programme national de traduction assistée par ordinateur de 1985 sous l'égide du Centre canadien de Recherche sur l'Informatisation du Travail. Quant au Japon, c'est le projet annoncé en 1982 à l'ICOT (Institute for new Generation Computer Technology) soutenu par le MITI (Japan's Ministry of International Trade and Industry), qui a donné une réelle dynamique. Ce vaste projet avait comme objectif d'asseoir les bases d'une société future fondée sur l'information. Il comprenait le développement de systèmes interactifs de traduction automatique assistée par des humains, de stations de travail pour les traducteurs, de systèmes de visualisation multilingues, de banques de données terminologiques, de systèmes de synthèse vocale et de reconnaissance de la parole pour l'élaboration de « téléphones traducteurs ». Ce projet, coordonné par l'Université de Kyoto, annonçait déjà la « troisième » et même la « quatrième » génération de systèmes de traduction.

Depuis 1990, nous assistons à un renouveau de l'intérêt pour la TA aux Etats-Unis pour des raisons économiques, notamment la montée des concurrents japonais et chinois, et la baisse des coûts de l'informatique. Cette période est caractérisée surtout par l'utilisation de l'approche empirique avec des traitements statistiques de grands corpus. Plusieurs systèmes de traductions automatiques tels que GoogleTranslate et LanguageWeaver sont construits sur des mémoires de traduction utilisant un très grand corpus bilingue aligné (*Canadian Hansards*⁵²). La mise en place de systèmes statistiques réputés être à bas coût est une fausse idée. Ce type de système engendre des coûts relatifs à la construction d'un très grand corpus parallèle et l'établissement d'un nombre de calculs conséquent. De plus, reposant sur des modèles purement statistiques, ils produisent des traductions peu fiables, voire même des contresens graves.

⁵² Corpus anglais-français des actes du parlement canadien.

A l'heure actuelle, force est de constater que les systèmes de traduction automatique ont fait beaucoup de progrès au niveau de la vitesse de traduction, mais peu au niveau de la qualité. La qualité des traductions fournies par les machines reste bien inférieure à la qualité des traductions humaines.

3. Traduction automatique et traduction assistée

La traduction automatique implique, au sens strict du terme, qu'il n'y ait aucune intervention humaine, en opposition à la traduction assistée par ordinateur (TAO). Or à l'heure actuelle, il n'existe pas de système de traduction de ce type, ainsi la dénomination « assistée par ordinateur » a gagné du terrain sur l'expression « traduction automatique », même si la frontière entre les deux domaines reste floue.

Selon la proportion du travail qui incombe à l'homme et à la machine, on distingue deux types de traductions assistées :

- le premier est la traduction automatique assistée par l'humain, la traduction est effectuée par l'ordinateur avec l'aide de l'homme.
- la seconde est la traduction humaine assistée par ordinateur. L'humain traduit en faisant appel à un ordinateur pour obtenir de l'aide comme l'accès à des bases de données terminologiques, la recherche de termes dans des dictionnaires ou encore la consultation de corpus bilingue.

De plus, les systèmes se différencient selon le moment de l'intervention de l'humain sur la machine. Certains systèmes dits de « traduction assistée » ont besoin de l'humain au cours de la traduction, il doit prendre certaines décisions (désambiguïsation d'un terme, choix de structure, etc.). C'est le cas des systèmes interactifs qui, face à des difficultés de traduction, stoppent la traduction pour interroger l'humain.

D'autres systèmes ne nécessitent d'intervention humaine qu'en amont (« pré-édition » du texte d'entrée) et en aval (« post-édition »). La machine effectue seule le processus de traduction du texte d'entrée vers celui de sortie, on parle de « traduction automatique ».

4. Les différentes architectures des systèmes de traduction automatique

4.1. Approche directe et approche indirecte

Les systèmes qui choisissent l'approche « directe » ne sont conçus que pour la traduction d'un couple de langues donné : par exemple, la première version de SYSTRAN fonctionnait exclusivement pour le couple russe-anglais. Ils réalisent un travail réduit à son minimum nécessaire pour les deux langues et ne font pas de distinction entre les tâches d'analyses et celles de génération. L'approche directe, avec une vision locale du texte, était un des traits caractéristiques des systèmes de « première génération ». Cependant une traduction mot à mot se heurte à de très nombreux problèmes tels que : l'homographie, la polysémie, la délimitation des groupes, problèmes de rattachement syntaxiques, etc.

Les systèmes, plus évolués, optent pour une approche « indirecte ». Ils accomplissent le passage du texte d'entrée (langue source) au texte de sortie (langue cible) par l'intermédiaire de représentations de ces textes. Ils séparent donc la tâche d'analyse du texte en langue source de la tâche de génération du texte en langue cible, ainsi chacun des deux modules d'analyse et de génération peut servir à nouveau quand le système est étendu à de nouvelles langues.

4.2. Approche par pivot et approche par transfert

L'approche indirecte peut elle même se diviser en deux méthodologies, on parlera :

- d'une approche dite par pivot,
- d'une approche dite par transfert.

4.2.1. Approche par pivot

Un système adoptant l'approche par pivot se compose de seulement deux modules :

- un module d'analyse produisant une représentation du texte d'entrée en langue source dans un langage pivot supposé indépendant de toute langue (ou au moins des deux langues considérées, source et cible),
- un module de génération qui construit à partir de la représentation précédemment produite un texte de sortie en langue cible.

Le pivot repose sur une analyse du texte source qui aboutit à une représentation intermédiaire indépendante des contraintes d'expression liées à la langue source et à la langue cible. Le système recourt à des primitives sémantiques, logiques, ou une représentation du monde pour construire cette représentation. La génération en langue cible se fait à partir de la représentation intermédiaire.

L'approche par pivot des systèmes de « deuxième génération » tels que le système russe-français du CETA de Grenoble semblait particulièrement pertinente dans le cas de deux langues typologiquement éloignées et dans les cas de traductions multilingues. En effet, le pivot est sensé être indépendant des langues. Ce système nécessite des analyses très profondes et fines quasi inaccessibles qui, pour des raisons pratiques de faisabilité, a été largement abandonné dans le milieu de la traduction automatique.

4.2.2. Approche à transfert

Comparativement avec un système à pivot, un système qui adopte une approche par transfert se compose de trois modules :

- un module d'analyse du texte en langue source,
- un module de génération du texte en langue cible,
- un module intermédiaire, en plus, dit de transfert qui transforme la représentation du texte R (obtenue en sortie de l'analyse) en représentation R' du texte (avant production de la génération)

Dans leur principe, les systèmes à transfert sont plus réalistes que les systèmes à pivot. Il s'agit dans ce type de système de traiter la chaîne de la langue source morphologiquement, afin de produire une représentation syntaxique de surface. Cette représentation est transformée par la suite en une représentation abstraite en éliminant tous les phénomènes sans incidence sur la traduction automatique. En d'autres termes on cherche à établir des règles de transfert, générales et abstraites tout en gardant un œil sur les cas particuliers. Le module de transfert convertit ensuite la représentation abstraite de la langue source dans une représentation abstraite de la langue cible. La génération utilise cette représentation abstraite de la langue cible pour générer une représentation syntaxique de surface et ensuite une chaîne cible. Les systèmes de « troisième génération », dont les représentations deviennent plus « profondes » tendent à rejoindre les systèmes à pivot, par conséquent, la divergence entre ces deux types de systèmes tend donc à s'estomper.

5. Conclusion

La traduction automatique a été depuis ses débuts jalonnée de toute une succession d'évolution allant de pair avec les progrès techniques. Les industriels ont pris le relais de l'armée et du monde politique et se sont ensuite efforcés de trouver des systèmes de traductions idéaux améliorant la qualité et la fiabilité de la traduction automatique. Dans cette évolution, l'approche empirique a vu le jour avec l'utilisation de modèles probabilistes, mais à quel prix ? Les utilisateurs de Google Translate en ont sûrement déjà fait la triste expérience. Les mauvaises traductions sont courantes et les risques de contresens graves ne sont pas à exclure. Quelles seraient alors les conséquences si les traductions concernaient des domaines sensibles tels que celui de la sécurité. De plus, ces systèmes sont réputés être des systèmes rigides et sans traçabilité, contrairement aux systèmes de traduction à base de règles.

Le système de traduction TACCT que nous proposons, en état de prototype, est un système de traduction contrôlée français-arabe de transfert à base de règles linguistiques. Ce système est le fruit d'un long travail de recherche et de réflexion sur les

domaines de traduction automatique et les langues contrôlées. Le modèle linguistique du système TACCT tire sa force des analyses microsystemiques menées à la fois sur les langues source et cible, d'un contrôle formel ciblé et d'une logique d'isomorphisme dans la construction des équivalences macrostructurales et microstructurales. Ce faisant, le système fonctionne également sur d'autres langues à savoir le chinois, l'anglais, le thaï et le russe et en cours d'extension sur l'espagnol et l'italien.

Chapitre 6

SYSTEMIQUE ET SYSTEME

1. Introduction

Le chapitre suivant portera comme son intitulé l'indique sur l'approche microsystemique. Nous commencerons par faire un bref retour dans le temps pour comprendre quand, comment et dans quelles conditions cette approche a vu le jour. Nous aborderons ensuite les premiers travaux en linguistique systémique menés par le centre Tesnière à travers la théorie SyGULAC. Ensuite s'ensuivront plusieurs questions de fond, quelle adaptation de l'approche SyGULAC pouvons-nous faire pour faciliter le passage du français vers l'arabe ? Quels sont les avantages que cette théorie présente pour le domaine de la traduction automatique ?, autant de questions auxquelles nous tenterons de répondre.

La traduction automatique, les expériences réalisées dans ce domaine le confirment, se heurte à divers problèmes dus dans la majorité des cas à la complexité du langage humain. Cette complexité génère des phénomènes linguistiques difficilement formalisables et auxquels la machine ne peut faire face sans des règles de transfert bien précises. Il est essentiel donc de se pencher sur la façon d'assainir la langue pour qu'elle devienne un objet prêt à l'automatisation. Les difficultés de la traduction automatique peuvent être rangées en deux catégories principales : les difficultés de modélisation et les difficultés de liaisons.

2. Naissance de la systémique

L'histoire de la systémique, de sa naissance à son avènement, se situe principalement au cours de la seconde moitié du vingtième siècle. Les spécialistes des différents domaines sont tous unanimes sur le fait qu'il faut réorienter la science, évolution des technologies oblige. La systémique émerge donc sous la forme d'une méthode nouvelle favorisée par ces avancées dans le domaine des sciences et des technologies auquel elle va en retour grandement contribuer. Au terme d'une certaine maturation que l'on peut qualifier de «technique», la théorisation de ce qui caractérise

cette nouvelle approche permet à la systémique de s'imposer, au delà d'un ensemble de méthodologies, en tant que véritable modèle.

La systémique a vu le jour aux Etats-Unis au début des années 50 grâce à la contribution de plusieurs chercheurs et grâce aussi aux résultats des recherches qu'ils ont menées dans différents domaines. L'approche systémique doit sa création à un professeur de mathématiques du Massachusetts Institut of Technologie (M.I.T.), spécialiste de la cybernétique, Norbert Wiener, à un neurophysiologiste Warren Mac Culloch, fondateur de la bionique « science qui affronte les inventions humaines et les processus biologiques (détection, locomotion, orientation), en vue d'offrir aux ingénieurs des modèles dont l'imitation peut être utile, et, à l'inverse, de mieux interpréter certains organes et fonctions par comparaison avec des inventions humaines », au biologiste Ludwing Von Bertalanffy ainsi qu'à l'électronicien Joy Forrester, professeur de management au MIT.

3. Voyage dans le temps, bref historique

Durant trois décennies, le dialogue constructif entre différentes disciplines scientifiques donne naissance aux grands principes qui seront les fondements de l'approche systémique. Le berceau de la systémique est sans doute le Massachusetts Institute of Technology à Boston, siège de ce dialogue. La circulation d'idées contribue au transfert de méthodes et de terminologies d'une discipline à l'autre. En trois bonds d'environ dix ans chacun, les travaux réalisés au M.I.T. vont amener au développement de la systémique et de la cybernétique.

Le point de départ de cette épopée est à situer dans le lien entre le mathématicien Norbert Wiener, professeur au MIT, et le neurophysiologiste Arturo Rosenblueth, chercheur à la Harvard Medical School. Chacun d'eux avait une petite équipe de chercheurs travaillant sous son égide à l'étude des similitudes pouvant exister entre le comportement des organismes vivants et celui des servomécanismes. Une des expériences célèbres dans l'histoire de la systémique établit une analogie troublante

entre une « maladie » qui touche les appareils de prédiction de la trajectoire d'avions et une maladie consécutive à des lésions du cervelet chez l'homme. Les servomécanismes doivent prévoir la future trajectoire d'un avion en prenant en compte des données de trajectoires sciemment enregistrées. Au cours de ce travail, deux éléments intriguent les deux chercheurs : l'intelligence du système, puisque à partir de données déjà enregistrées il pouvait prédire la future trajectoire de l'avion et des failles dans ce même système, car il entre dans un cycle d'oscillations incontrôlables quand les deux chercheurs ont essayé de réduire les frictions. Chez les humains le même phénomène existe, des lésions au niveau du cervelet conduisent à une amplification des mouvements et le patient ne peut plus par exemple porter un verre d'eau à sa bouche. En effet, dans les deux cas, un phénomène oscillatoire vient perturber la stabilité du système. Cette analogie a permis la mise en évidence de l'existence de boucles de rétroaction dans les systèmes aussi bien physiologiques que techniques.

Une série de dix séminaires est organisée entre 1946 et 1953. Au milieu de ce bouillonnement d'idées se constituent les bases d'un langage commun qui deviendra celui de la systémique. Norbert Wiener publie en 1948 son célèbre ouvrage *Cybernetics*. Dans la même année, Claude Shannon, ingénieur des télécommunications, publie la théorie mathématique de l'information et fonde la théorie de l'information.

Les années 50 voient arriver un outil qui permet de traiter la complexité organisée, l'ordinateur. Le plus rapide Wirlwind II fut conçu au MIT en 1951 fonctionnant pour la première fois d'ailleurs avec une mémoire magnétique ultra-rapide inventée par un jeune ingénieur Jay W. Forrester, électronicien au laboratoire des servomécanismes. Jay Forrester à la tête du Lincoln laboratory est un ingénieur chargé par l'Air Force de coordonner la mise au point d'un système d'alerte et de défense avec radars et ordinateurs dans le but de détecter et d'empêcher les attaques ennemies sur le sol américain. Cette expérience va d'ailleurs largement contribuer à accroître la conscience de Forrester sur l'importance de l'approche systémique dans la compréhension, la conception et le contrôle d'organisations complexes faisant intervenir hommes et machines interconnectés en temps réel, c'est-à-dire capables de prendre des décisions vitales au fur et à mesure de l'arrivée des informations. Forrester va s'intéresser, à

travers cette nouvelle perception, au sein de la Sloan School of Management du MIT, à l'organisation de l'entreprise conçue comme système complexe.

La société pour l'étude des systèmes généraux (Society for general systems research) créée en 1954 par le biologiste Ludwig von Bertalanffy tente d'englober la cybernétique dans un ensemble plus vaste qui prendra en charge l'étude générale des systèmes.

Dès les années 70, un certain nombre de conditions culturelles, scientifiques et institutionnelles vont permettre, à partir des approches préalablement développées, de constituer une réelle science des systèmes. Le passage de science utilisant et développant des approches systémiques à la systémique à proprement parler est dû à une interfécondation des idées entre plusieurs disciplines et également au rôle déterminant de quelques chercheurs principalement américains et français. Ils ont d'ores et déjà pris conscience de la nécessité d'une synthèse à un niveau théorique de l'ensemble des lois fondant la science des systèmes. On peut par exemple citer : H. Simon, H. von Foerster, J. Forrester, E. Morin, I. Prigogine, H. Atlan, J.L. Le Moigne, etc.

Le premier à s'être lancé dans cette lourde tâche est le biologiste Ludwig Von Bertalanffy, en 1968, à New York, il rassemble l'ensemble de ces travaux dans le livre *General System Theory* traduit sous le titre : *La théorie générale des systèmes*. Il développe dans cet ouvrage, qui fait la part belle aux systèmes biologiques, un certain nombre de concepts tels que ceux des systèmes ouverts ou encore d'équifinalité, etc. Il est le premier à prôner une appréhension globale du système. Il insiste sur l'importance des relations entre les différents éléments constitutifs du système et sur l'importance de la compréhension de ces relations. Il rompt ainsi avec la pensée classique d'une saisie analytique des éléments du système.

Quasiment au même moment, Herbert Simon et Kenneth Boulding contribuent également à la théorisation des principes développés dans le cadre de leurs propres travaux. La définition des concepts d'arborescence et de niveaux d'organisations composent le premier palier vers un effort de construction d'une typologie des systèmes qu'ils proposent à huit niveaux. Le premier niveau correspond aux objets statiques et

simples de la physique et de la chimie, le dernier niveau est celui de la socio-culture. De bas en haut, on avance dans une complexification croissante.

La tentative de généralisation des différents travaux par Jay Forrester aboutit à la constitution d'une discipline nouvelle appelée dynamique des systèmes. Le caractère opérationnel de la dynamique des systèmes explique pour grande part le succès immédiat qu'elle connaît dans de nombreux domaines tels que la biologie, l'environnement ou encore la gestion. Les difficultés auxquelles se heurte cette dynamique, surtout en termes de quantification et de modélisation, conduisent à la différenciation de la systémique en deux courants : celle de première génération et celle de deuxième génération.

On trouve ainsi la systémique de première génération directement liée à la cybernétique. Elle s'appuie sur des méthodes quantitatives et sur l'outil informatique pour, au-delà de la « simple » compréhension, essayer de prévoir son comportement.

La systémique de deuxième génération, également nommée System Thinking, s'inscrit dans une perspective quelque peu différente de celle de première génération. Dans cette vision de la systémique, il faut mettre l'accent sur l'intelligibilité du comportement du système. Son but est de concevoir des modèles qualificatifs permettant d'entrer dans l'intelligence du phénomène et d'en orienter plus ou moins l'action. Cette systémique de seconde génération a connu également un vif succès principalement au Mexique mais aussi très largement en France. Jean-Louis Le Moigne, professeur à l'université d'Aix/Marseille a largement apporté sa contribution au développement et aux applications de la pensée systémique, avec par exemple l'animation du groupe MCX (Modélisation de la complexité).

Il convient de citer aussi l'AFSCET, association française de sciences des systèmes Cybernétiques, cognitifs et Techniques, qui travaille en groupes d'étude pluridisciplinaires facilitant les échanges entre chercheurs et co-organisant régulièrement un congrès européen.

4. Raisons d'une apparition

Voici près d'un demi-siècle que l'approche systémique a vu le jour pour aider les chercheurs à affronter la complexité des situations que la seule approche cartésienne, qui isole les éléments, ne permettait pas d'appréhender. La systémique est apparue dans un contexte où l'approche analytique est remise en cause, affaiblie par la complexité et l'incertitude de notre époque. La société et les technologies modernes sont devenues si complexes que les moyens traditionnels ne sont plus en mesure d'y répondre, et le besoin de moyens holistiques et systémiques s'est fait de plus en plus sentir.

Les systèmes réclament une intervention à tous les niveaux, c'est ce que nous verrons d'ailleurs dans la suite de ce chapitre, avec le système de la langue où le problème au cours de la traduction automatique est un problème posé principalement par un nombre de variables en interrelation. L'objectif est donc d'envisager parmi diverses possibilités, au milieu d'un réseau extrêmement emmêlé d'interactions, celles qui permettront une traduction fiable avec le maximum d'efficacité.

Le regard systémique propose un nouveau départ avec des moyens d'action nouveaux et ingénieux en vue de débloquent les situations de paralysie. L'objectif est donc d'amorcer une dynamique, dans un cadre de pensée opératoire, pour passer d'une situation de flou à une situation de clarté. Cette nouvelle pensée donnera des moyens opérationnels qui permettront à la fois de comprendre la réalité telle qu'elle est, agir sur elle mais aussi de la considérer autrement et de changer de stratégie d'action.

Le point de vue systémique a pénétré un grand nombre de domaines scientifiques et technologiques y compris celui de la langue avec la théorie SyGULAC (Voir chap6/6.1. Théorie SyGULAC).

Plusieurs facteurs peuvent justifier l'apparition et le développement de la systémique comme approche révolutionnaire, cependant deux facteurs pourraient être considérés comme principaux :

- la reconnaissance de la complexité des objets et des phénomènes qui nous entourent et la complexité aussi des problèmes qui affectent les différents domaines de recherche, d'organisation, et de gestion. Nous parlons souvent de problèmes pour désigner des difficultés à résoudre ou des situations de paralysie auxquelles une méthode doit faire face. Dans notre cas, les problèmes sont des problèmes linguistiques liés à la traduction automatique.
- la prise de conscience de l'existence de propriétés communes à tous les systèmes mais aussi de méthodes générales identiques applicables à la majorité des systèmes quelle que soit leur nature. L'analyse détaillée des systèmes révèle qu'un fond conceptuel unitaire existe entre les systèmes, loin d'être des analogies superficielles, ce sont des isomorphismes sur des principes généraux de représentation, de modélisation et de contrôle qui font surface.

5. Définition d'un système

La notion de système a fait l'objet de plusieurs définitions étant donné qu'il constitue la matière première de la systémique. Certes, les disciplines dans lesquelles les systèmes ont été définis et l'orientation idéologique de chaque auteur ont considérablement influencé les définitions attribuées au système, Il n'en demeure pas moins qu'une base commune peut être établie.

Un système est pour Joël de Rosnay (1975) : « *un ensemble d'éléments en interaction dynamique, organisé en fonction d'un but* » mettant ainsi l'accent sur la finalité d'un système. Morin (1991) quant à lui considère le système comme « *une unité globale organisée d'interrelations entre éléments, actions ou individus.* » insistant de ce fait sur le caractère global du système, et les interactions qui caractérisent les éléments qui le composent.

Enfin, vient la définition de Ferdinand de Saussure du système qui est particulièrement bien articulée. Elle fait surtout surgir, en la mettant en lien à celles de totalité et d'interrelation, le concept d'organisation : le système est « *une totalité* »

organisée, faite d'éléments solidaires ne pouvant être définis que les uns par rapport aux autres en fonction de leur place dans cette totalité» (Saussure, 1916).

De toutes ces définitions, émergent trois notions fortes et principales qui caractérisent un système :

1. la complexité,
2. la globalité,
3. l'interaction.

Ces notions qui se dégagent de la notion de système sont des notions complémentaires et en aucun cas contradictoires. Elles constituent les propriétés présomptives de la systémicité. Elles font partie du domaine de la systématique qui cherche à montrer la complexité d'une organisation, et à explorer ses diverses composantes.

5.1. La complexité

La naissance de la systématique est liée avec une prise de conscience, notamment dans le domaine scientifique, de la complexité croissante du monde. Plus le monde devient complexe, plus l'homme prend conscience de sa complexité. Des systèmes simples de l'univers tels que l'atome, nous sommes passés à des systèmes beaucoup plus élaborés comme le mammifère ou encore les sociétés. Cette organisation laissant souvent apparaître une architecture interne dont la base est identique.

Cette caractéristique explique que l'organisation au premier abord est une simplification, le phénomène de redondance propre à l'organisation conduit à une diminution de la variété, cela fige l'objet. Dans un deuxième temps, l'organisation génère des possibilités relationnelles nouvelles entre sous-ensembles organisés. La tendance va donc s'inverser et la complexification va augmenter. L'apparition de sous-ensembles permet l'introduction de niveaux d'organisation (théorisés notamment par Le Moigne).

Chapitre 7

Modèle syntaxico- sémantique du système TACCT

1. Introduction

L'analyse syntaxique d'une langue est plus difficile que ce que nous pensons au premier abord. Chaque langue possède son propre agencement syntaxique et construit ses structures syntaxiques d'une façon spécifique qu'il n'est pas facile de traduire. La performance d'un système de traduction repose en grande partie sur le traitement syntaxique qui constitue l'ossature du système. La traduction automatique ne peut se faire que si l'on résout le problème du passage du modèle syntaxique source au modèle syntaxique cible. La sémantique occupe aussi une place importante dans le traitement des langues naturelles car l'usage de la langue par l'homme à des fins diverses engendre une multiplication des sens d'un même mot, d'une phrase, etc. La langue n'est qu'un reflet de notre vécu, elle véhicule la pensée humaine qu'elle concrétise dans des formes graphiques : « *En fait, à chaque langue correspond une organisation particulière des données de l'expérience [...]. Une langue est un instrument de communication selon lequel l'expérience humaine s'analyse différemment dans chaque communauté* ». (Martinet, 1970)

De ce fait, maîtriser cette sémantique si souple n'est pas une tâche facile et son appréhension devient parfois pour les concepteurs des traducteurs automatiques un tracas sans fin. Comment résoudre la complexité sémantique d'une langue lorsqu'on sait que même les traducteurs humains en font leur cheval de bataille « *traduire la langue ou transmettre le sens, c'est l'objet même de la traduction qui est en cause dans cette alternative* ». (Seleskovitch, Lederer, 1986)

Le défi est ensuite d'arriver à une approche globale capable de faire le lien entre le niveau syntaxique et le niveau sémantique. Cette approche sera la base de chaque transfert que le système effectuera.

Le système de traduction automatique contrôlée TACCT relève ce défi en répondant à des exigences bien précises : rapidité, fiabilité mais surtout traçabilité. Sa conception repose sur une approche multidimensionnelle en associant plusieurs paramètres : analyse intralangue et interlangue, analyse microsystemique, contrôles statique et dynamique, syntaxe lexicalisée et règles de transfert.

Au niveau de la syntaxe, compte tenu de l'importance des représentations syntaxiques, le premier objectif dans notre système était d'arriver à un isomorphisme dans les structures recensées entre le français et l'arabe. Nous avons de ce fait procédé premièrement à l'analyse interne et externe des textes de sécurité afin d'identifier les récurrences syntaxiques et contrôler ensuite les textes dans le but de rapprocher les deux langues syntaxiquement soit par un parallélisme parfait avec une syntaxe symétrique, soit avec des structures éloignées et une syntaxe de ce fait asymétrique généralement lexicalisée. L'analyse syntaxique s'est également appuyée sur la théorie SyGULAC pour avoir un modèle en microsystemes facilement manipulable (arguments, compléments, attributs, négation, etc.).

La sémantique dans le système TACCT est mise au même niveau que la syntaxe. En effet, le système dispose d'un modèle syntaxique si puissant qu'il est capable d'assurer le transfert sémantique et de détecter certaines valeurs sémantiques bien avant la phase de transfert. Nous parlons à ce niveau précis du modèle syntaxico-sémantique TACCT et de macrostructures miroir.

2. Notion de macrostructure miroir

La notion de macrostructure miroir telle qu'elle est établie dans le système TACCT est une notion nouvelle à la fois par la démarche et la réflexion menée pour sa mise en place et par le résultat lui-même. Elle tire son originalité du modèle de représentation syntaxico-sémantique qui prend en compte les divergences et les convergences entre les deux langues source et cible mais aussi des règles de contrôle qui s'y rapportent. Ce mariage entre méthode comparative et contrôle a pour effet de ramener les deux langues sur un plan plus au moins isomorphe en fixant des structures argumentales stables et équivalentes. Lorsque les écarts sont importants au moment de la traduction des textes non contrôlés, le contrôle par ses différents types (nominalisation, contrôle statique, contrôle dynamique) intervient par un ensemble de transformations considérées comme autant de règles de contrôle, pour essentiellement désambiguïser et rapprocher les deux langues. Le résultat final de ce contrôle et du travail autour du

CONCLUSION

GENERALE

Dans une situation d'urgence, la communication écrite via des systèmes informatisés devient un enjeu majeur et un moyen incontournable pour la gestion de crise. C'est dans ce contexte que notre recherche prend toute sa place et revêt toute son importance. La communication écrite aujourd'hui s'avère être à l'origine de véritables catastrophes quand cette dernière est émaillée d'erreurs et d'ambiguïtés. La langue naturelle est un système de communication complexe truffée d'ambiguïtés sur tous les plans lexical, syntaxique, morphologique et phonétique générant des difficultés de compréhension, des approximations et des malentendus entre individus. Un message émis dans une situation d'urgence mal écrit ou mal formulé peut donc constituer un réel danger pour les populations notamment dans des domaines sensibles tels que la sécurité. La communication comme moyen de gestion de crise devient alors un vrai casse-tête pour les professionnels de la sécurité qui font face non seulement à la complexité des langues en situation monolingue mais d'autant plus en situation multilingue. La traduction des textes en lien avec la sécurité peut se révéler une opération à haut risque qui peut mettre en danger les vies humaines, en témoigne l'erreur de traduction dans l'affaire des sur-irradiés d'Epinal.

Notre recherche compte tenu du contexte actuel, constitue une avancée incontestable dans le domaine de la sécurité en apportant une approche nouvelle dans le domaine des langues contrôlées et de la traduction automatique. Nous avons pu au terme de cette recherche minutieuse et structurée apporter une réponse concrète aux problèmes de la communication écrite dans les domaines à sécurité critique en établissant des normes de rédaction rigoureuses adaptées aux besoins des professionnels de la sécurité. Nous avons au cours de notre travail mis en exergue les caractéristiques linguistiques des protocoles de sécurité, défini leur mode de fonctionnement et identifié les failles qu'ils présentent dans leur fiabilité communicative. Cela nous a permis d'agir directement sur ces textes afin d'améliorer leur lisibilité et leur compréhensibilité avec un dispositif de contrôle formel monolingue et bilingue pour la paire des langues français et arabe.

Ce dispositif de contrôle introduit des notions nouvelles dans le domaine des langues contrôlées qui constituent autant d'outils normatifs de contrôle. Ainsi nous

avons identifié dans notre recherche, la nominalisation comme outil normatif de contrôle bilingue. Nous avons également mis en lumière au cours de notre recherche deux types de contrôles complémentaires, le contrôle statique et le contrôle dynamique. Une langue contrôlée à notre sens, dans une perspective de traduction automatique, doit prendre en compte les particularités linguistiques de chaque langue séparément et les particularités communes aux deux langues. Les règles de contrôle deviennent par conséquent un compromis entre deux systèmes linguistiques différents qui les rapproche tout en préservant les particularités de chacun.

Ce faisant, nous avons réussi à mettre en place un système de traduction automatique contrôlée à base de règles linguistiques TACCT. Le système TACCT repose sur un modèle de transfert syntaxico-sémantique isomorphique qui assure des traductions fiables et rapides du français vers l'arabe. La traduction automatique devient alors dans une situation de crise et de bilinguisme un moyen de communication efficace et non plus une opération hasardeuse et approximative.

Au cœur de ce modèle linguistique, nous avons introduit le concept nouveau des macrostructures miroir construites autour du verbe et d'un ensemble d'éléments obligatoires et facultatifs.

L'application de la microsystemique à travers la théorie SyGULAC, a permis de travailler sur de petits microsystèmes et d'établir des règles de transfert précises assurant un passage rapide du français vers l'arabe. Elle prouve par ailleurs qu'il est possible d'obtenir un isomorphisme entre deux langues à priori éloignées.

Cette recherche novatrice dans les domaines des langues contrôlées et de la traduction automatique français-arabe ouvre des perspectives de recherche dans d'autres domaines du traitement automatique des langues. Le modèle linguistique du système TACCT peut aisément faire l'objet d'une adaptation pour d'autres applications du TAL du fait de la précision des normes établies et de la rigueur de ses formalisations. De surcroît, le système de traduction TACCT français-arabe qui fonctionne aussi pour l'anglais, le chinois et le thaï a été élargi pour le russe et est en cours d'élargissement pour l'italien et l'espagnol.

REFERENCES ET BIBLIOGRAPHIE

1. Références

ADAM, Jean-Michel, (1987), *Types de séquences textuelles élémentaires*, Pratiques, no 56, Metz, pp. 54-79.

BALLY, Charles, (1965), *Linguistique générale et linguistique française*, Bernes.

BEDDAR, Mohand., CARDEY, Sylviane, (2008), *Comment contrôler les langues pour la traduction automatique : l'exemple du français pour l'arabe*, in revue RML6, 2008, Acte du Colloque International en Traductologie et TAL, 7 et 8 juin 2008, Oran, Algérie.

BEDDAR, Mohand, (2009), *French to Arabic Machine Translation Isomorphic Syntax, Use of Terminal Sequences*, in ISMTCL Proceedings, International Review Bulag, PUFC, ISSN 0758 6787, ISBN 978-2-84867-261-8, pp 38-42, Besançon.

BEDDAR, Mohand, (2010), *Approche Systémique et Système de Traduction Automatique TACCT Contrôlée Français-Arabe Appliqué au Protocoles de Sécurité*, in Colloque international en traductologie et TAL, 17-18 janvier 2010, Oran, Algérie.

BEDDAR, Mohand, CARDEY, Sylviane, (2010), *Système TACCT Contrôlée Français-Arabe Appliqué au Protocoles de Sécurité*, colloque international Langues, cultures et professionnalisation dans un contexte mondialisé. Affaires et Traduction dans l'espace euro-arabe, 17-18 avril 2010, Le Caire, Égypte.

BEDDAR, Mohand, (2010), *Traduction Automatique des Syntagmes Nominaux (N1-de-N2) du français vers l'arabe Dans les Protocoles de Sécurité*, IIIème édition de la Conférence internationale LEXIQUE COMMUN / LEXIQUE SPÉCIALISÉ, 8- 9 septembre 2010, Galați, Roumanie.

BEDDAR, Mohand, (2011), *Syntaxe Comparative et Langues Contrôlées en vue de la Traduction Automatique des Alertes et des Protocoles de Sécurité du Français vers l'Arabe Système TACCT Contrôlée Français-Arabe*, actes des journées d'études toulousaines Jétou 2011, 7-8 avril 2011, Toulouse, France.

BEDDAR, Mohand, (2013), *Approche Systémique et Contrôle Intra et Inter-langues dans le Système de Traduction Automatique TACCT Contrôlée Français-Arabe Appliquée aux Protocoles de Sécurité*, colloque international TRALOGY II, Trouver le sens : où sont nos manques et nos besoins respectifs ?, sous la direction de Nicolas Froeliger, Joseph Mariani, Jean-François Nominé, Alain Wallon, Dominique Durand-Fleischer et Mojca Bozic, 17-18 janvier 2013, Paris, en cours de publication.

BEDDAR, Mohand, (2013), *Intra- and Inter- Language Control for French to Arabic Machine Translation in the Security Domain*, KEMCS 2013, sera publié dans *Advanced Materials Research* (ISSN:1022-6680), March 3-4, 2013, Phuket, Thailand.

BENVENISTE, Émile, (1966), *Formes nouvelles de composition nominale*, Bulletin de la Société de linguistique de Paris.

BIBER, D., (1988), *Variations across speech and writing*, Cambridge University Press, Cambridge.

CABRE, Maria Teresa, (1998), *La terminologie – Théorie, méthode et applications*, Les Presses de l'université d'Ottawa, Armand Colin.

CARDEY, S., (1987), *Traitement algorithmique de la grammaire normative du français pour une utilisation automatique et didactique*, thèse d'état sous la direction de M. le Professeur Y.GENTILHOMME, Université de Franche-Comté, Besançon, pages 1 à 52.

CARDEY, S., GREENFIELD P., (2004), *Systemic Language Analysis and Processing*, in *Proceedings of the International Conference on Language, Brain and Computation*, 3-5 October 2002, Venice, Italy, Benjamins, 2004.

CARDEY, S., BEDDAR, M., GREENFIELD, P., JIN G., MIKATI, Z., RENAHY, J., SEKUNDA, G., (2010), *Le projet LiSe, Linguistique, normes, traitement automatique des langues et sécurité : du data et sense mining aux langues contrôlées*, in *Actes du WISG 2010, Workshop Interdisciplinaire sur la Sécurité Globale*, Université de Technologie de Troyes, 26 & 27 Janvier 2010, 10 pages, CD-ROM.

CHARAUDEAU, Patrick, (1992), *Grammaire du sens et de l'expression*, Paris, Hachette Éducation.

CHAROLLES, M., (1997), *L'encadrement du discours : Univers, champs, domaines et espaces*, Cahier de Recherche Linguistique 6, Université de Nancy 2.

DALADIER, Anne, (1978) *Problèmes d'analyse d'un type de nominalisation en français et de certains groupes nominaux complexes*, thèse de 3e cycle, L.A.D.L., Université de Paris VII.

DARMESTER, A., (1877), *De la création actuelle des mots nouveaux dans la langue française et des lois qui la régissent*, thèse de doctorat, Paris.

DE ROSNAY, Joël, (1975), *Le macroscopie. Vers une vision globale*, Paris, Edition Seuil.

DEPECKER, Loïc, (2002), *Entre signe et concept : éléments de terminologie générale*, Presses de la Sorbonne nouvelle, Paris, 196 p.

DIXON, R.M.W., (1991), *A new Approach to English Grammar, on Semantic Principles*, Oxford University Press.

DUBOIS, J., (1969), *Énoncé et énonciation*, *Langages*, 13, pp100-110.

DUBOIS, Jean et coll., (1994), *Dictionnaire de linguistique et des sciences du langage*, Larousse, Paris.

EURIN, S. et HENAO, M., (1992), *Pratiques du français scientifique*, UREF, Paris: Hachette F.L.E.

FUCHS, Catherine, (1996), *Les ambiguïtés du français*, Éditions Ophrys, Coll. L'essentiel français.

GALISSION, Robert et COSTE, Daniel, (1976), *Dictionnaire de didactique des langues*, Hachette, Paris.

GANIER, F., (1999), *Le traitement cognitif: Déterminant de la conception de documents procéduraux*, Thèse, Université de Bourgogne (Dijon).

GARCIA-DEBANC, C., (2001), *L'étude des discours procéduraux aujourd'hui : travaux linguistiques et psycholinguistiques*, Langages, 141, 3-9.

GENELOT, Dominique, (1992), *Manager dans la complexité, réflexions à l'usage des dirigeants*, INSEP.

GIRY-SCHNEIDER, Jacqueline, (1987), *Les prédicats nominaux en français. Les phrases simples à verbe support*, Volume 18 de Langue & cultures, Éditeur Droz, Paris.

GROSS, Gaston, (1996), *Les expressions figées en français, nom composés et autres locutions*, Paris : Ophrys, 161 p.

GROSS, G. et GUENTHENER, F., (2002), *Comment décrire une langue de spécialité ?*, Dans Cahiers de lexicologie, (2002-1), revue internationale de lexicologie et lexicographie, publiés par l'Institut de linguistique française (CNRS), N°80. Paris : Honoré Champion. pp. 179-199.

GROSS, Maurice, (1998), *La fonction sémantique des verbes supports*, Travaux de linguistique, no. 37.

GUESPIN, Louis, (1991), *La circulation terminologique et les rapports science, technique, production*, dans GAUDIN et ASSAL, pp. 59-79.

GUILBERT, Louis, (1973), *La spécificité du terme scientifique et technique*, dans Langue française. N° 17, Larousse, Paris, pp 5-18.

GUILBERT, Louis, (1975), *La créativité lexicale*, Librairie Larousse, Paris, 285 p.

GUIRAUD, Pierre, (1971), *La sémiologie*, PUF, coll. « Que sais-je ? ».

HARRIS, Z., (1988), *Language and information*. New York: Columbia University Press.

HEURLEY, L., (1994), *Traitement de textes procéduraux. Étude de psycholinguistique cognitive des processus de production et de compréhension chez les adultes non experts*, Thèse de Doctorat de l'Université de Bourgogne, Dijon, sous la direction de Jean-Emile Gombert et Michel Fayol.

HEURLEY, L., (1997), *Vers une définition du concept de texte procédural : le point de vue de la psycholinguistique*, Cahiers du Français Contemporain, 4, pp 109-133.

HEURLEY, L., (2001), *Compréhension et utilisation de textes procéduraux : l'effet de l'ordre de mention des informations*, Revue Française de Linguistique Appliquée 2001/2, Volume VI, pp. 29-46.

HOFFMANN, L., (1979), *Towards a theory of LSP. Elements of a methodology of LSP analysis*, Fachsprache 79(1-2), pp 12-17.

HUNSTON, S., FRANCIS G., (2000), *Pattern Grammar. A Corpus-driven Approach to the Lexical Grammar of English*, Amsterdam and Philadelphia, Benjamins.

JAKOBSON, R., (1963), *Essais de linguistique générale*, Paris, Editions de Minuit.

JESPERSEN, Otto, (1942), *A Modern English Grammar on Historical Principles*, Part VI, Copenhagen: Ejnar Munksgaard.

KOCOUREK, Rostislav, (1991), *La langue française de la technique et de la science, vers une linguistique de la langue savante*, Wiesbaden, Oscar Brandstetter Verlag.

KOHLER, C., KÉKENBOSCH, C., VERSTIGGEL, J.C., (2000), *La compréhension d'un texte procédural : un processus à profondeur variable*, Journal International de Psychologie, 35(6), 258-269.

LERAT, Pierre, (1995), *Les langues spécialisées*, PUF, Paris.

LONGACRE, R.E., (1982), *Discourse typology in relation to language typology*, In Text processing: Text analysis and generation, text typology and attribution, Sture Allén

edition, pages 457-86, Proceedings of the Nobel Symposium 51. Stockholm: Almqvist and Wiksell.

LUDWIG VON BERTALANFFY, K., (1980), *Théorie générale des systèmes*, Bordas-Dunod.

LÜGER, Heinz-Helmut, (1995), *Pressesprache*, Tübingen, Niemeyer.

MARESCHAL, G., (1989), *Repérage d'unités terminologiques dans le contexte de l'enseignement de la traduction spécialisée*, Meta, vol. 34, n. 3, Les Presses de l'Université de Montréal.

MARTINET, André, (1970), *Éléments de linguistique générale*, Paris, Armand Colin.

MARTINOT, Claire, (1997), *Les verbes supports dans l'acquisition de la syntaxe*, Actes du Colloque International de Besançon sur l'acquisition de la syntaxe en langue maternelle et en langue étrangère, Annales littéraires de l'Université de Franche-Comté, 631, Paris, Les Belles Lettres, pp 87-106.

MOLINO, J., (1990), *Thèses sur le langage, le discours, la littérature et le symbolique*, Zeitschrift für französische Sprache und Literatur, K.W. Hempfer et P. Blumenthal, Franz Steiner Verlag, Stuttgart.

MORTARIA GARAVELLI, Bice, (1988), *Typologia des testi*, Lexikon der romanistischen Linguistik, vol IV, Holtus & al. Edition, Tübingen, Niemeyer, pp 157-168.

MOUNIN, G. (1979), *La linguistique comme science auxiliaire dans les disciplines juridiques*, Meta, vol. 24, n°1, pp 9-17.

NAZARENKO, A., (2004), *Donner accès au contenu des documents textuels. Acquisition de connaissances et analyse de corpus spécialisés*, Mémoire d'Habilitation à Diriger des Recherches de l'Université Paris 13. Villetaneuse.

PHAL, A. (1971), *Vocabulaire général d'orientation scientifique (V.G.O.S.) - Part du lexique commun dans l'expression scientifique*, Paris, Didier.

POTTIER, Bernard, (1992), *Sémantique générale*, Paris, Presses Universitaires de France.

QUAMAR, Hassan, (1996), *Quand dire, c'est : écrire-comment-faire. Un autre type de texte le Recettal*, Thèse soutenue à l'Université Lumière, Lyon II.

QUEMADA, B., (1978), *Technique et langage*, Dans Gille, B. (ed.), *Histoire des techniques*, Gallimard, Paris.

RASTIER, F., (1989), *Sens et textualité*, Paris, Hachette, Collection HU Linguistique.

RASTIER, F., (2001), *Arts et sciences du texte*, Presse Universitaire Française, Paris.

RIEGEL, Martin, PELLAT, Jean-Christophe & RIOUL, René, (1994), *Grammaire méthodique du français*, Presses Universitaires de France, Paris.

SAGER, J.-C., DUNGWORTH, D. and MCDONALD, P., (1980), *English Special Languages. Principles and practice in science and technology*, Wiesbaden: Brandstetter.

SAGER, J. C., (1994), *Language Engineering and Translation: Consequences of automation*, Amsterdam, John Benjamins.

SAUSSURE, F., de (1972 [1916]), *Cours de linguistique générale*, édition critique préparée par Tullio De Mauro, Paris : Payot. p. 13-43, 155-184.

SELESKOVITCH, Danica et LEDERER, Marianne, (1986), *Interpréter pour traduire*, Didier érudition, Paris.

SERRES, Michel, (1980), *Hermès IV, Passage du Nord-Ouest*, Edition de Minuit.

SINCLAIR, J.M., (1991), *Corpus Concordance Collocation*, Oxford, Oxford University Press.

SPAGGIARI, L., & al, (2009), *From Controlled Language to Controlled expression*, in ISMTCL Proceedings, International Review BULAG, PUFC, ISSN 0758 6787, ISBN 978-2-84867-261-8, pp 38-42, Besançon.

VARANTOLA, K., (1986), *Special Language and General Language: Linguistic and Didactic Aspects*, ALSSED-LSP Newsletter 9, 2 (23). 10-19.

VERGELY, P., (2008), *Communications professionnelles et usage de la langue naturelle : une question de risque(s) ?*, In revue Enjeux de l'information et de la communication.

VINAY, J.P. et J., DARBELNET, (1969), *Stylistique comparée du français et de l'anglais*, Paris: Didier.

VIVES, Robert, (1983), *Avoir, prendre, perdre: constructions à verbe support et extensions aspectuelles*, Thèse de troisième cycle, LADL, Université Paris VII.

VIVES, Robert, (1984), *L'Aspect dans les constructions nominales prédicatives: avoir, prendre, verbe support et extension aspectuelle*, In *Linguisticae Investigationes* 3 (1).

VIVES, Robert, (1993), *La prédication nominale et l'analyse par verbes supports*, L'information grammaticale, n° 59, Paris.

WERLICH, E., (1975), *Typologie der Texte*, Heidelberg, Quelle & Meyer.

WIERZBICKA, A., (1982), *Why Can You Have a Drink When You Can't *have an eat?*, *Language*, 58, 4, pp 753-799.

WILLEMS, D., (1969), *Analyse des critères d'auxiliarité en français moderne*, Travaux de linguistique, no. 1.

2. Bibliographie

ABDELOUAFI, Ghenimi, (2006), *Conception d'un modèle de traduction arabe-français appliqué au domaine des mathématiques*, thèse, Centre Tesnière, université de Franche-Comté.

ABI AAD, Albert, (2001), *Le système verbal de l'arabe comparé au français*, Paris, Maisonneuve et Larose.

ACL/EACL, (2001), *Workshop, Arabic Language processing: status and Prospects*, Toulouse, <http://www.elsnet.org/acl2001-arabic.html>.

ADAM, Jean-Michel, (1985), *Quels types de textes?*, *Le français dans le monde*, no 192, pp. 39-44.

ADAM, Jean-Michel, (1992), *Les textes, types et prototypes*, Paris, Nathan.

ADAM, Jean-Michel, (2001), *Types de textes ou genres de discours ? Comment classer les textes qui disent de et comment faire ?* *Langages*, 141, 10-27.

AECMA, (1989), *A guide for the preparation of aircraft maintenance documentation in the international aerospace maintenance language*, Association Européenne des Constructeurs de Matériel Aérospatial, Paris.

AFNOR, (1990), *Terminologie*, Norme ISO 1087, Paris.

ALLERTON, David, (2002), *Stretched Verb Constructions in English*, London : Routledge.

ARNOLD, D.J., (2003), *Why translation is difficult for computers*. In Harold Somers, editor, *Computers and Translation: A translator's guide*, pages 119–142. John Benjamins, Amsterdam.

BADAWI, E., CARTER, M. et GULLY, A., (2004), *Modern Written Arabic. A comprehensive Grammar*, London and New York, Routledge, collection Comprehensive Grammars.

BAILLY, F, GODARD, O., (1995), *Revue Internationale de Systémique*, Dunod Afcet, Volume 9, numéro 4.

BEACCO, J. C., & MOIRAND, S., (1995), *Les enjeux des discours spécialisés*, Paris, Presses de la Sorbonne nouvelle.

BOGACKI, Krzysztof, (2009), *Controlled languages and Machine Translation*, ISMTCL Proceedings, International Review Bulag, PUFC, ISSN 0758 6787, ISBN 978-2-84867-261-8, p. 49-55.

BOUCHON-MEUNIER, B., TIJUS, Bernard, (1997), *Revue Internationale de Systémique*, Dunod Afcet, Volume 11, numéro 1.

BOUHADIBA, Farouk, (2003), *Traduction et structures syndétiques en arabe: le cas de la particule waw*, BULAG N° 28, France.

BOUHADIBA, Farouk, (2004), *L'infixation en Arabe et la concaténation en TA : questionnement d'un linguiste*, 2^{èmes} Journées sur la Traductologie, 5-6 mai 2004, Centre Lucien Tesnière, Besançon, France.

BOUHADIBA, Farouk, (2004), *Le parler arabe d'Oran : de la ruralité vers l'urbanité*, Conférence Internationale : Evolutions des parlers urbains arabes, l'impact des mouvements migratoires et des changements sociaux, 20-23 octobre 2004, Aix en Provence, France.

BOUHADIBA, Farouk, (2005), *Langues et cultures, systèmes et traduction*, Colloque META 7-9 avril 2005, Journal des traducteurs, vol 50, n°4 décembre 2005. Revue META, CD ROM (Montréal). <http://www.erudit.org/livre/meta/2005/000200co.pdf> (Article commun sur la traduction automatique de langues d'apparentements génétiques différents. Article pour la Revue Canadienne META (février 2005).

BOUHADIBA, Farouk, (2007), *La langue arabe et le TAL : étude de cas*, RML 5, 2007, Actes du 1^{er} Colloque International en Traductologie et TAL, 9-11 avril 2007, Université d'Oran, Algérie.

BOUHADIBA, Farouk, (2011), *Prospects for a Trilingual Electronic Dictionary in the Field of Petroleum and Hydrocarbons (Revisited)*, TICLA – ICTAL, Technologies d'Information et de Communication en Langue Arabe, Paris 23-24 juin 2011 – IUFM de Gennevilliers, France.

BOUTET, Josiane, (1995), *Paroles au travail*, l'Harmattan collection langage et travail.

BURIDANT, P., KLEIBER, G. et PELLAT, J.-C. (éds),. (2001), *Par monts et par vaux. Itinéraires linguistiques et grammaticaux*, Mélanges de linguistique offerts au Professeur Martin Riegel pour son 60e anniversaire par ses collègues et amis, Leuven, Editions Peeters.

BUVET, Pierre-André, (1995), *Levées d'ambiguïté et domaines d'emplois*, IVes Journées Internationales LTT, octobre 1995, Lyon.

BUVET, Pierre-André, (1996), *Procédures de filtrage des composés N de N*, 15e Colloque Lexique-Grammaire des Langues Romanes, octobre 1996, Munich.

BUVET, Pierre-André, (2000), *Les groupes nominaux LE N MODIF à MODIF obligatoire*, 19e Colloque Lexique-Grammaire des Langues Romanes, septembre 2000, Ericeira.

BUVET, Pierre-André, (2006), *La description des verbes supports stylistiques dans un dictionnaire de prédicats*, Colloque La marque lexicographique : quel avenir ? novembre 2006, Chypre.

BUVET, Pierre-André, (2009), *Les représentations des structures prédicat-argument : Présentation*, Langages 176, Paris : Armand Colin, pp. 63-79, [en collaboration avec Xavier BLANCO].

BUVET, Pierre-André, (2010), *Les séquences adjectivales du type en N*, III Jorndas Intercianoles de Phraseologia Constrativa, novembre 2010, Alicante.

CARDEY, S., GREENFIELD, P., (2002) *Systemic Language Analysis and Processing*, In actes de la Conférence internationale sur Language, Brain and Computation, 3-5 Octobre 2002, Venise, Italie.

CARDEY, S., (2003), *Modélisation, systémique, traductibilité*, BULAG, 28, Université de Franche-Comté, Besançon.

CARDEY, S., GREENFIELD, P., (2005), *A Core Model of Systemic Linguistic Analysis*, In Proceedings of the International Conference RANLP-2005 Recent Advances in Natural Language Processing, Borovets, Bulgaria, 21-23 September 2005, pp. 134-138.

CARDEY, S., GREENFIELD, P., (2008), *Micro-systemic linguistic analysis and software engineering: a synthesis*, in revue RML6, 2008, Acte du Colloque International en Traductologie et TAL, Oran.

CARDEY, S., GREENFIELD, P., ANANTALAPOCHAI, R., BEDDAR, M., DEVITRE, D., AND JIN, G., (2008), *Modelling of Multiple Target Machine Translation of Controlled Languages Based on Language Norms and Divergences*, Proceedings of ISUC2008 (Second International Symposium on Universal Communication) published by the IEEE Computer Society, Osaka, Japan, pp 322-329.

CARDEY, Sylviane, (2009), *Machine Translation of Controlled Languages for More Reliable Human Communication in Safety Critical Applications*, The 11th International Symposium on Social Communication, ACTAS, Santiago de Cuba, p. 330-335.

CHAIRET, Mohamed, (1996), *Fonctionnement du système verbal en arabe et en français*, Orphys, Paris.

CHAROLLES, M., (1980), *Les formes directes et indirectes de l'argumentation*, Pratiques, n°28, Metz.

CHAROLLES, M., (1988), *Les plans d'organisation textuelle: périodes, chaînes, portées et séquences*, Pratiques, 57, 3-14.

CHOLEWA, Joanna, (2009), *Controlled Remarks Concerning Writing Texts in Controlled Polish*, ISMTCL Proceedings, International Review Bulag, PUFC, ISSN 0758 6787, ISBN 978-2-84867-261-8, p. 62-68.

COHEN, Marcel, (1924), *Le système verbal sémitique et l'expression du temps*, Paris, Leroux.

CONDAMINES, A., (2008), *Peut-on prévenir le risque langagier dans la communication écrite?* In P. Vergely (dir.) *Langage et Société, Communiquer au travail en situation de risque*, 2008/3, n° 125, 77-97.

DALADIER, Anne, (1996), *Le rôle des verbes supports dans un système de conjugaison nominale et l'existence d'une voix nominale en français*, *Langages*, 121.

DE ROSNAY, Joël, (1975), *Théorie des systèmes et communication*, *Le macroscopie : vers une vision globale*. Paris. Le Seuil.

DELISLE, Jean, (1993), *La traduction raisonnée*, Ottawa, Presses de l'Université d'Ottawa.

DONNADIEU, Gérard & KARSKY, Michel, (2002), *La systémique: penser et agir dans la complexité*, *Liaisons*.

DURAND, D., (1979), *La systémique*, Paris, P.U.F., Coll. « Que sais-je ? », n°1795.

EL-DAHDAH, Antoine, (1991), *Dictionnaire de la grammaire arabe universelle, arabe-français*, Librairie du Liban, 1991.

ENGLEBERT, A., (1998), *L'infinitif dit de narration*, éd. Duculot, Paris.

EQUOY HUTIN, Séverine & al., (2009), *Valeurs et Enjeux des écrits au travail*, SEMEN 28.

EUGENE, J., (1981), *Aspects de la théorie générale des systèmes*, *Recherches interdisciplinaires*, Maloine éditeur, Paris.

EURIN BALMET, S., HENAO DE LEGGE, M., (1992), *Pratiques du français scientifique*, Hachette, Paris.

FLUHR, Christian, (1977), *Algorithmes à apprentissage et traitement automatique des langues*, Thèse d'état N°1863. Université de Paris-Sud. Orsay, 15 juin 1977.

GAVIEIRO-VILLATTE, E., SPAGGIARI, L., (1999), *Open ended overview of controlled language*, in BULAG n°24, Génie Linguistique et Génie Logiciel, Université de Franche-Comté, Besançon, pp. 89-100.

GAVIEIRO, E., (2001), *Vers un modèle d'élaboration de la terminologie d'une langue contrôlée ; application aux textes d'alarmes en aéronautique pour les futurs postes de pilotage*, Thèse de doctorat sous la direction de Sylviane Cardey, Centre de Recherche Lucien Tesnière, Université de Franche-Comté, Besançon.

GENTILHOMME, Yves, (1982), *Lecture d'un texte scientifique : introduction*, n° 35, octobre 1982, pp. 100-115.

GENTILHOMME, Yves, (1985), *Essai d'approche microsystemique, Théorie et pratique : Application dans le domaine des sciences du langage*, Berne, Francfort-s. Main, New York, Lang.

GENTILHOMME, Yves, (1994), *Termes et symboles discours hétérogènes. Quelques hypothèses sémiologiques*, in CLAS André et Pierrette BOUILLON (éds.), TA-TAO : Recherches de pointe et applications immédiates, AUPELFUREF.

GIFAS, (1990), *Guide du rédacteur*, Groupement des Industries Françaises Aéronautiques et Spatiales, Paris.

GIRAUD, S., (dir.), ROMAN, A., (1988), *L'infinitif : une approche comparative*, éd. Presses de l'Université de Lyon, Lyon.

GIRE, Alain, (1988), *Théorie ouverte des systèmes*, L'interdisciplinaire.

GLEDHILL, Chris, (2007), *La portée : seul dénominateur commun dans les constructions verbo-nominales*, In Frath, P. Pauchard, J. & Gledhill, C.(éds.) Actes du Premier Colloque Res per nomen 24-26 mai 2007. Reims: Presses de l'Université de Reims-Champagne-Ardenne.

GRAND'HENRY, J., (2000), *Grammaire arabe à l'usage des Arabes. Traduction française et commentaires des Éléments d'arabe, morphologie et syntaxe*, II, de Rachid Chartouni (Beyrouth), Série Pédagogique de l'Institut Linguistique de Louvain, vol. (24). Peeters.

GROSS, Maurice, (1986), *Grammaire transformationnelle du français, 1-Analyse du verbe*, Paris, Éditions Cantilène.

GROSS, Maurice, (1986), *Grammaire transformationnelle du français, 2- Syntaxe du nom*, Paris, Éditions Cantilène.

GUESPIN, Louis., (1993), *Normaliser ou standardiser?* Le langage et l'homme, 28, n°4, 213-222.

GUILBERT, Louis (en collaboration avec Jean PEYTARD), (1973), *Les vocabulaires techniques et scientifiques*, Langue française, n° 17, Larousse, Paris.

HABERT, B., NAZARENKO, A. & SALEM, A., (1997), *Les linguistiques de corpus*, Paris, Armand Colin, collection U.

HAKKAK, G. et NEYRENEUF, M., (1996), *Grammaire active de l'arabe littéral*, Paris, Le Livre de Poche.

HARRIS, B., (1988), *Bi-text, a new concept in Translation Theory*, Language Monthly, n° 54.

HARRIS, B., (1988), *Are you Bi-Textual ?*, Language Technology, n° 7.

HARRIS, Z., GOTTFRIED, M., RYCKMAN, T., MATTICK, Jr P., DALADIER, A., HARRIS, T., HARRIS, S., (1989), *The Form of Information in Science, Analysis of Immunology Sublanguage*, Dordrecht, Netherlands, Kluwer Academic Publisher.

HARRIS, Z., (1991), *A theory of language and information, A mathematical approach*, Oxford University Press.

HUTCHINS, W.J., (1979), *Linguistic models in machine translation*, UEA papers in linguistics 9, 29-52.

HUTCHINS, W. J., SOMERS, H. L., (1992), *An Introduction to Machine Translation*, Londres, Academic Press.

ISABELLE, Pierre et al, (1978), *TAUM-Aviation : description d'un système de traduction automatisée des manuels d'entretien en aéronautique*, Communication au congrès (COLING'78).

ISABELLE, Pierre, BOURBEAU, Laurent, (1984), *TAUM-AVIATION: its technical features and some experimental results*, Computational Linguistics, 11:1, 18-27.

KITTREDGE, R., (1982), *Variation and Homogeneity of Sublanguages*, Walter de Gruyter, Berlin.

KLEIBER, G., (1990), *La sémantique du prototype*, Paris, Ed. Presses Universitaires de France.

KLEIBER, G., (1999), *Problèmes de sémantique, La polysémie en questions*, Paris, Presses Universitaires du Septentrion.

LE GALLOU, F. & BOUCHON-MEUNIER, B., (1992), *Systémique Théorie et applications*, Tec et Doc Lavoisier.

LEMOIGNE, Jean-Louis, (1977), *La théorie du système général: Théorie de la modélisation*, Paris, Ed. Presses Universitaires de France.

LEMOIGNE, Jean-Louis, (1984), *La théorie du système général*, Paris, Ed. Presses Universitaires de France.

LEMOIGNE, Jean-Louis, (1990), *La modélisation des systèmes complexes*, Paris, Afcet systèmes, Dunod.

LERBET, Georges, (1993), *Approche systémique et production de savoir, Alternances développements*, L'harmattan.

L'HOMME, M.C., (2004), *La terminologie : Principes et Techniques*, Presses Universitaires de Montréal.

LONGACRE, R.E., (1970), *Sentence structure as a statement calculus*, *Language*, 46, 783-815.

MARTINET, André, (1999), *Le syntème*, In *La Linguistique*, 35 (2). 11-16.

MEJRI, S., (1993), *Séquences figées et expression de l'intensité. Essai de description sémantique*, *Cahiers de lexicologie* 65, pp. 111-122.

MEL'CUK, I., (1994), *Cours de morphologie générale*, PU de Montréal, vol. 2, Montréal.

MORIN, Edgar, (1991), *La méthode*, Œuvre complète, Ed. Seuil.

MORIN, Edgar, (2005), *Introduction à la pensée complexe*, Edition du Seuil, Paris.

MOURLHON-DALLIES, F., (2008), *Enseigner une langue à des fins professionnelles*, Didier, Paris.

NOAILLY, M., (1990), *Le substantif épithète*, PUF, Paris, Coll. Linguistique Nouvelle.

PEIRCE, Charles S., (1978), *Ecrits sur le signe*, traduits et commentés par Gérard Deledalle, Edition du Seuil, Paris.

PHAL, A., (1966), *Langue scientifique et analyse linguistique*, in *L'Initiation des étudiants et chercheurs étrangers à la langue scientifique et technique*, CRÉDIF, Saint-Cloud, p. 27 et sqq.

PICHT, H., and J., DRASKAU, (1985), *Terminology: An Introduction*, England, The University of Surrey.

REY, A., (1979), *La terminologie, noms et notions*, col. Que sais-je ? PUF, Paris.

ROMAN, André, (1983), *Étude de la phonologie et de la morphologie de la koine arabe*, 2 vols, Aix-en-Provence, Université de Provence.

SATO, S., NAGAO, M., (1990), *Toward Memory-based Translation*. In Proceedings of the 13th International Conference on Computational Linguistics, COLING-90, Helsinki.

SATO, S., (1992), *CTM: An Example-based Translation Aid System*, In Proceedings of the 14th International Conference on Computational Linguistics, COLING-92, Nantes.

SATO, S., (1995), *MBT2: A Method for Combining Fragments of Examples in Example-Based Machine Translation*, Artificial Intelligence, n° 75.

SEMMAR, Nasredine, ELKATEB-GARA, Faiza, and FLUHR, Christian , (2005), *Using a stemmer in a natural language processing system to treat Arabic for cross-language information retrieval*. In Proceedings of the Fifth Conference on Language Engineering, pages 1–10, Cairo, Egypt.

SPAGGIARI, L., (2002), *Formalisation des rapports syntaxiques d'une langue contrôlée : application aux messages d'alerte dans le poste de pilotage avion*, Thèse de doctorat, Université de Poitiers.

SPAGGIARI, L., BEAUJARD, F., and E., CANNESSON, (2005), *A Controlled Language at Airbus*, Lingvisticae Investigationes Tome 28 fascicule 1, John Benjamins, Amsterdam, pp 107-122.

TESNIERE, Lucien, (1959), *Eléments de syntaxe structurale*, Klincksieck, Paris.

TOLAS, J., (2004), *Le français pour les sciences*, Grenoble, PUG.

VAUQUOIS, B., (1992), *La Traduction Automatique*, In Pottier, B. (Ed.), Les Sciences du langage en France au XX^{ème} siècle, Paris, Peeters.

VENDRYES, Joseph, (1968), *Le langage. Introduction linguistique à l'histoire*, Paris, Albin Michel.

VIKNER, C., (1980), *L'infinitif et le syntagme infinitif*, Revue romane, Vol. 15, Numéro 2, pp. 252-291.

VIRBEL, J., (1985), *Langage et métalangage dans le texte du point de vue de l'édition en informatique textuelle*, in Cahiers de Grammaire, 10, pp. 5-72.

YATCHINOVSKY, Arlette, (1999), *L'approche systémique pour gérer l'incertitude et la complexité*, ESF Editeur, Collection Formation Permanente, Paris.

3. Sites internet

MUEGGE, Uwe. *A controlled language with ten rules*. Disponible sur : <http://www.muegge.cc/controlled-language.htm>

Projet « MESSAGE. Alert messages and protocols ». Ressources. Disponible sur : <http://message-project.univ-fcomte.fr/resources-en.html>

<http://www.inf.ed.ac.uk/teaching/courses/mt/lectures/history.pdf>

<http://www.afscet.asso.fr/SystemicApproach.pdf>

<http://www.almaany.com/home.php?language=arabic>

<http://www.lexilogos.com/>

<http://www.reefnet.gov.sy/education/kafaf/index.html>

ANNEXES

Corpus